The Dissertation Committee for Hua Zhang Wise Certifies that this is the approved version of the following dissertation:

Cellulose Biosynthesis in *Physcomitrella patens*

Committee:

R. Malcolm Brown, Jr., Supervisor

G. Barrie Kitto

David L. Herrin

Alan M. Lloyd

Stanley J. Roux, Jr.

Cellulose Biosynthesis in *Physcomitrella patens*

by

Hua Zhang Wise, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2007

Dedication

This dissertation is dedicated to my husband, Allen.

Acknowledgements

I will always be indebted to Dr. R. Malcolm Brown Jr. for accepting me into his lab, for having faith in me, for his guidance and constant encouragement. I would also like to thank each of the committee members: Dr. David L. Herrin, Dr. G. Barrie Kitto, Dr. Alan M. Lloyd, and Dr. Stanley J. Roux Jr. for their invaluable suggestions and instructions. Last but not least, I would like to acknowledge the support of all the colleagues in Dr. R. Malcolm Brown, Jr.'s laboratory (especially Dr. David R. Nobles, Jr. and Dr. Inder M. Saxena), without whose support I never would have made it this far, and with a special tip of the hat to Richard Santos.

Cellulose Biosynthesis in *Physcomitrella patens*

Publication No._____

Hua Zhang Wise, Ph.D.

The University of Texas at Austin, 2007

Supervisor: R. Malcolm Brown, Jr.

Physcomitrella patens has become a model system to study plant biology. 8 cellulose synthase (CesA) genes were identified by searching against *Physcomitrella* EST database. Two of these genes, PpCesA6 and PpCesA7 are the first full-length CesAs to be identified. These two genes are highly similar to each other, both on the cDNA and genomic DNA levels. They both have 13 introns and 12 exons. The first introns are more than 1kb. The proteins they encode both have 1096 amino acids. There are only three amino acid differences in the proteins they encode. PpCesA6 and PpCesA7 share 74% amino acid identity with Monterey pine (*Pinus radiate*) PrCesA10, 72% amino acid identity with quaking aspen (*Populus tremuloide*) PtrCesA6, 71% amino acid identity with maize (*Zea mays*) ZmCesA7 and three rice (*Oryza sativa*) CesAs, 65%-68% amino acid identity with *Arabidopsis* CesAs. The deduced proteins of PpCesA6 and PpCesA7 contain the D, D, D, QXXRW motif in the form of DDG, DCD, TED, QVLRW, which is the catalytic region of cellulose synthases. Two other pairs of CesA genes, PpCesA3 and

PpCesA8, PpCesA4 and PpCesA10, also show high similarity. PpCesA2 and PpCesA9 are pseudogenes.

By taking advantage of the high efficiency homologous recombination in *Physcomitrella* nuclear DNA, a C-terminus GFP fusion construct was produced for PpCesA6. Expression analysis showed that PpCesA6 is expressed in both protonemata and young gametophore. In protonemata, PpCesA6 is expressed in both chloronema and caulonema cells, but not in every cell. In young gametophore, PpCesA6 is expressed in axillary hairs and rhizoids. Confocal miscrocopy study shows that PpCesA protein is localized on the plasma membrane and it is randomly dispersed.

The gene targeted knockout constructs of PpCesA6 and PpCesA7 were produced. The null mutants of PpCesA6 and PpCesA7 single knockout as well as double knockout were generated by the PEG (polyethylene glycol)-mediated protoplast transformation. Both single knockout mutants did not show obvious phenotypic differences from the wild type. The double knockout mutants had reduced stem length. The stem lengths of the wild type, PpCesA6 knockout mutant, PpCesA7 knockout mutant and double knockout mutant growing on BCD and BCDAT media were 3.93±0.45mm and 3.51±0.08mm, 3.82±0.46mm and 3.5±0.3mm, 3.65±0.68mm and 3.73±0.49mm, 2.75±0.22mm and 2.65±0.43mm, respectively.

A cellulose synthase-like C gene (CslC4) was identified by searching against the *Physcomitrella* EST and genomic DNA databases. The protein it encodes is 694 amino acids. The D, D, D, QXXRW motif is in the form of DDS, DAD, VED, QQHRW. PpCslC4 genomic DNA has 4 small introns in the coding region. There is also one small

intron at the 5'-UTR. The deduced PpCslC4 protein shows 72% similarity with PpCslC2 and PpCslC3, 65% similarity with PpCslC1. When compared with other organisms, PpCslC4 protein shows more than 60% similarity with *Arabidopsis* and *Oryza sativa* CslC proteins.

A gene targeted knockout construct was produced for PpCslC4. The null mutants were generated by the PEG-mediated protoplast transformation. PpCslC4 mutant did not show any obvious phenotypic differences from the wild type.

Table of contents

List of Tables

List of Figures

Chapter 1 Introduction

Cellulose is the most abundant biopolymer on earth (Brown, 2004). Cellulose is found in plants, algae, the slime mold *Dictyostelium*, bacteria and tunicates (Blanton *et al.*, 2000; Brown, 1985; Matthysse, 2004; Ross *et al*., 1991). Cellulose is the major component of the plant, algae and *Dictyostelium* cell wall. In higher plants, the cell wall is a complicated network of cellulose, hemicellulose, glycosylated proteins, pectins and lignin. Although the molecular approach has been widely used to study cellulose biosynthesis, there are relatively many discoveries to be made in this area.

A number of plant systems have been used for studying cellulose biosynthesis such as the cotton fiber, *Arabidopsis*, maize and poplar. The moss *Physcomitrella patens* has become a model system in the study of plant biology (Schaefer and Zryd, 2001; Schaefer, 2002). Compared with higher plants, *Physcomitrella patens* has a dominant haploid gametophyte in its life cycle. This feature makes mutant isolation and genetic analysis much simpler than seed plant systems. Another unique feature about this moss is that it has a high frequency of homologous recombination in its nuclear DNA. This makes targeted knockout more precise.

The goals of my research are:

(1) Identify and characterize cellulose synthase and cellulose synthase-like genes in *Physcomitrella patens*;

(2) Study the functions of these genes by targeted gene knockout;

(3) Use *Physcomitrella patens* to establish a new system for the study of cellulose biosynthesis.

1.1 Cellulose

Cellulose is a β-1, 4-linked glucose polymer with the glucose residues oriented 180^0 to each other. The linked monomers together form the glucan chain. Aggregates of these chains constitute a microfibril. Cellulose microfibrils are about 3nm thick and contain 36 glucan chains (Brown and Saxena, 2000). Celluose microfibrils have a crystalline structure. The multiple hydroxyl groups on the glucose residues form both intra- and inter-molecular hydrogen bonds. The hydrogen bonds only have a small contribution to the crystalline structure. The Van der Waals bonds are the major forces that bind the glucan chains together in the microfibril (French *et al.*, 1993; Delmer, 1999). Together the hydrogen bonds and Van der Waals forces stabilize the microfibrils and make them insoluble, giving them resistance to chemical degradation and making them mechanically strong.

Cellulose provides strength and structure for the cell wall. Cellulose microfibrils are the scaffolds for binding other cell wall components. Many plant cells have both a primary cell wall and a secondary cell wall. The primary cell wall forms during cell expansion. The secondary cell wall develops inside the primary wall after the cell has stopped growing. The cellulose microfibrils in the primary cell walls are arranged in an irregular network while in the secondary cell walls they are deposited in a parallel orientation (Carpita and Gibeaut, 1993; Franz and Blaschek, 1990).

Most crystalline cellulose produced in nature is cellulose I. There are two sub-allomorphs of cellulose I, namely I_α and I_β. I_α is found in bacteria and many algae whereas I_β is found in plants (Sugiyama *et al.*, 1991). I_α and I_β have a different crystal structure and hydrogen bonding pattern (Nishiyama *et al.*, 2003). In cellulose I allomorph, the glucan chains are parallel. Glucan chains can also be anti-parallel. This leads to the formation of the cellulose II allomorph. Most cellulose II allomorphs are products of dissolution and re-precipitation and they rarely exist in nature (Delmer, 1999). Cellulose I can be converted to cellulose II under strong alkaline conditions.

1.2 Cellulose biosynthesis and cellulose synthase genes

Roelofsen and Preston predicted that cellulose is synthesized on the cell membrane by multienzyme complexes (Preston 1974, Roelofsen 1958). The cellulose synthase complexes (called terminal complexes-TCs) were first discovered by Brown and Montezinos on the plasma membrane in *Oocystis apiculata* in 1976 (Brown and Montezinos, 1976). Since then, TCs have been found in *Acetobacter* (Brown *et al.*, 1976), numerous algae (Tsekos, 1999), the slime mold *Dictyostelium* (Blanton, 2000), and higher plants (Muller and Brown, 1980). The TCs are arranged in either linear or 6-fold symmetries, the latter complexes called rosette terminal complexes (Muller and Brown, 1980). Linear TCs have been found in *Acetobacter*, *Dictyostelium*, tunicates and some algae. Higher plants, bryophytes and some green algae have the six-fold symmetry rosette TCs (Brown, 1990; Tsekos, 1999).

Although the studies on cellulose biosynthesis started in the 1920s, it was not until 1990 that the first cellulose synthase gene was identified by Saxena *et al.* in *Acetobateria xylium* (Saxena *et al.*,1990). This success led to the identification of the cellulose synthase genes in plants. The first plant cellulose synthase genes were identified by sequencing a cotton fiber cDNA library. Amino acid sequences derived from two cDNA clones from this library showed sequence similarity to the bacterial cellulose synthase. Further study showed that the protein expressed in *E. coli* from one of these cDNAs bound with the cellulose synthase substrate UDP-glucose (Pear *et al.*, 1996). Since then, cellulose synthase genes have been identified in more than 250 plant species (http://cellwall.stanford.edu). Cellulose synthase genes have been studied intensively in *Arabidopsis*, rice and poplar. There are 10 cellulose synthase genes in *Arabidopsis* (Richmond, 2000; Richmond and Somerville, 2000), at least 12 in maize (Appenzeller *et al.*, 2004), at least 10 in rice (Tanaka *et al.*, 2003) and at least 7 in aspen (Joshi *et al.*, 2004).

Cellulose synthase proteins are type 2 glycosyl transferases (http://afmb.cnrs-mrs.fr/~pedro/CAZY/gtf_2.html). They share a conserved structure. Plant cellulose synthase proteins are transmembrane proteins with 8 transmembrane domains, two in the amino-terminus and six in the carboxyl-terminus. The amino-terminal portion of the protein contains a Zinc finger domain. The Zinc domain is thought to be involved in protein-protein interaction. Following the amino-terminus are the 2 transmembrane domains. Between the amino-terminus and carboxyl-terminus is the globular domain. The amino-terminus, carboxyl-terminus, and the globular regions face the cytosol. The

globular domain contains a D, D, D, QXXRW motif. Q is glutamine, R is arginine, W is tryptophan, and X is any amino acid. The aspartic acid residues (D) are variably spaced, and these residues are thought to bind with the UDP-glucose substrate. These residues are probably involved in catalysis. The QXXRW motif is also thought to be part of the catalytic site, although what role this motif plays in catalysis is not clear at present. The globular domain of cellulose synthases is more hydrophilic. It has 4 conserved sub-regions, U1-U4. The 3 aspartic acid residues are in the U1, U2, U3 regions and the QXXRW motif is in the U4 region. Compared with bacterial cellulose synthase proteins, higher plant cellulose synthase proteins have 2 plant-inserted regions. The first one is a highly conserved plant region (CR-P) that is located between U1 and U2. The other one is a class specific region (formerly called hypervariable region -HVR) that is located between U2 and U3 (Pear *et al.*, 1996; Vergara and Carpita, 2001).

The role of cellulose synthase involved in cellulose biosynthesis was further proved by reverse genetics and immunochemistry. The first cellulose-deficient temperature-sensitive mutant rsw1 (*AtCesA1*) showed reduced cellulose content, accumulated non-crystalline β-1, 4-glucan, and disassembled rossette structure (Arioli *et al.*, 1998). The rsw1 mutant provided the first molecular proof for the role of AtCesA1 in cellulose biosynthesis. The first direct proof for the existence of cellulose synthase complexes came when an antibody produced from the catalytic region of cotton cellulose synthase gene successfully labeled rosette TCs in the plasma membrane (Kimura *et al.*, 1999).

Among the 10 cellulose synthase genes in *Arabidopsis*, *At*CesA1, 2, 3, 5, 6 are involved in primary cell wall biosynthesis. *At*CesA1 and *At*CesA3 are gametophytic lethal while the *At*CesA6 null mutant shows a relatively mild phenotype. A recent study has shown that *At*CesA5 and *At*CesA2 are partially redundant with *At*CesA6, and this evidence suggested that they most likely compete with *At*CesA6 for the same binding site in the cellulose synthase complex (Burn *et al.*, 2002; Desprez *et al.*, 2002, 2007; Persson *et al.*, 2007; Somerville, 2006). *At*CesA4, *At* CesA7 and *At*CesA8 are involved in the secondary wall synthesis (Taylor *et al.*, 2000, 2003). So far there have been no reports on gene function redundancy of secondary wall biosynthesis.

Each particle in the six-fold symmetry rosette TCs is thought to have six CesA subunits, and it is believed that a total of 36 glucan chains make up the microfibrils (Delmer, 1999). Many questions arise from this statement, such as: 1) how many CesA proteins are present in the subunit? ; 2) how are these proteins arranged? ; 3) how do these proteins assemble microfibrils? So far, we can't answer all of these questions, except for only a few aspects. Evidence from *Arabidopsis* has shown that three separate cellulose synthase genes are required in the same cell at the same time during the secondary cell wall biosynthesis (Taylor *et al.*, 2003). Investigation of the cotton fiber cellulose synthase catalytic subunits indicated that dimerization of these subunits occurs via oxidation of the zinc-binding domains (Kurek *et al.*, 2002). A recent study on the *At*CesA7 suggests that phosphorylation of this protein may target it for degradation via a proteasome dependent pathway. This may be a mechanism that plants use to maintain the levels of cellulose synthase proteins (Taylor, 2007). Many facts still remain unknown

about cellulose, such as: the cellulose synthesis mechanism, the relation between the synthase complex (e.g. the rosette terminal complex in plants) and microfibrils, and how the synthesis of microfibrils and crystallization link together.

1.3 The cellulose synthase superfamily

There is a group of proteins that show structural similarity to cellulose synthases. They are all integral membrane proteins, and they also contain the D, D, D, QXXRW motif. The genes encoding these proteins are named "cellulose synthase-like (CSL)" genes (Cutler and Somerville, 1997; Saxena and Brown, 1995). Together with cellulose synthase genes, they belong to the same superfamily. It is believed that CSL genes are involved in the synthesis of non-cellulosic polysaccharides (Richmond and Somerville, 2001). CSL genes in *Arabidopsis* have been classified into six subfamilies, CslA, CslB, CslC, CslD, CslE, and CslG (Richmond and Somerville, 2000). Two additional subfamilies, CslF and CslH are found in rice (Hazen *et al.*, 2002). Together, there are 8 CSL subfamilies.

Unlike CesA genes which have a Zn-binding domain in the N-terminal region, all the CSL families lack this domain. The proteins encoded by CslD genes are as large as CesAs (around 1000 amino acids). The CslD family is approximately 45% similar to the CesA family at the amino acid level (Richmond and Somerville, 2000). All other CSL families encode smaller proteins. The *Arabidopsi*s CslA proteins are around 500 amino acids, CslC proteins are a little less than 700 amino acids, and CslB and CslG proteins are between 700 and 800 amino acids.

Among all the CSL families, the CslD family is the most similar to the CesA genes. CslD genes are thought to be the oldest in the cellulose synthase superfamily (Richmond and Somerville, 2001). CslA and CslC genes are more similar to bacterial cellulose synthase genes than to plants. In a phylogenetic tree, CslG is placed as ancestral to other members of the Csl/CesA clade (Nobles and Brown, 2004).

Understanding of the functions of CSL genes is very limited. Gene families with homologous sequences often have similar, but not identical functions. Cellulose synthases catalyze the formation of β-1, 4-linked glucose chains, and it is likely that CSL proteins catalyze the formation of other β-1, 4-linked cell wall components such as mannans, galactans, xyloglucans, glucomannans, and xylans (Richmond and Somerville, 2001). So far, we only have limited knowledge about some subfamilies of the CSL genes.

Expression of cellulose synthase-like genes in insect cells reveals that CslA family members encode mannan synthases (Dhugga *et al.*, 2004; Liepman, 2005). Analysis of expressed CslA proteins from diverse land plants, including angiosperms, gymnosperms and bryophytes in insect cells, shows that CslA protein catalyzes mannan and glucomannan synthesis in vitro (Liepman *et al.*, 2007).

A recent study showed that by screening a developing nasturtium seed cDNA library, the CslC gene is involved in the synthesis of the xyloglucan (XyG) backbone. A further study in *Arabidopsis* showed that *At*CslC4, which is highly similar to the nasturtium CslC gene, is coordinately expressed with other genes involved in XyG biosynthesis (Cocuron *et al.*, 2007).

The CslD gene in both *Arabidopsis* and rice is expressed in the root hair and is required for root hair elongation (Favery *et al.*, 2001; Kim *et al.*, 2007; Wang *et al.*, 2001). Subcellular localization of the $CslD_3$ gene in *Arabidopsis* showed that this gene was located on the endoplasmic reticulum, indicating that this gene was required for synthesis of a non-cellulosic wall polysaccharide (Favery *et al.*, 2001).

The expression of a rice CslF gene in *Arabidopsis* (which lacks this gene subfamily) has provided a direct evidence that CslF genes participate in (1, 3; 1, 4)-beta-d-glucan biosynthesis (Burton *et al.*, 2006). The functions of other Csl gene families remain unknown.

1.4 The moss *Physcomitrella patens*

The moss *Physcomitrella patens* is a non-vascular, multicellular bryophyte. It has become a model system to study plant biology (Cove *et al.*, 1997; Reski, 1998; Schaefer, 2001, 2002). Like other mosses, *Physcomitrella patens* is photoautotrophic with a life cycle dominated by a haploid gametophyte. The life cycle of *Physcomitrella* is shown in Figure 1-1. A germinating spore develops into hyphal-like tubular cells known as protonema. Further development of the protonema gives rise to the prothallium gametophyte. *Physcomitrella* is monoecious. Both antheridia and archegonia are produced on the same shoot. Spermatozoids (male gametes) need water to reach the archegonia to fertilize the egg cells (or oogonia). After fertilization, the zygote develops into a diploid sporophyte. The sporophyte is matrotropic and permanently attaches to the gametophyte. The sporophyte forms a spore capsule (sporangium), which undergoes

meiosis to produce numerous haploid spores and this completes the life cycle. One spore capsule contains about 4000 spores. The entire life cycle takes less than 3 months (Cove, 1992).

The most important feature about this moss is that its nuclear DNA has a very high efficiency of homologous recombination. *Physcomitrella* is the first land plant found that has a high efficiency of gene targeting similar to the yeast *Saccharomyces cerevisiae* (Schaefer and Zryd, 1997; Schaefer, 2001). In higher plants, the frequency of homologous recombination is only about 10^{-4}–10^{-5} (Britt and May, 2003) whereas in *Physcomitrella*, this frequency can be up to 100% (Kamisugi *et al.*, 2006). This facilitates the analyses of genes and their functions by targeted gene knockouts. This feature, in combination with the predominant haploid life cycle, makes *Physcomitrella* an ideal organism for genetic studies.

The predominant haploid stage of *Physcomitrella patens* has two developmental phases: the protonema and gametophore. The protonema phase can be further divided into chloronema and caulonema. The chloromena has a large number of chloroplasts and it has transverse cell walls. It shows one-dimensional growth. The caulonema has fewer chloroplasts than the chloromena, and its cell walls are oblique (figure 1-2). The caulonema shows three-dimensional growth. It undergoes differentiation and gives rise to multiple buds, which in turn develop into adult gametophytes. Differentiation of the caulonema involves major changes in the cell walls and cytoplasmic organization. Under certain culture conditions, the diploid stage can be maintained as a protonema. The protonema tissue undergoes very fast tip-growth. The apical cells of chloronemal

filaments can grow 2-5μm/h and divide every 22-26 hours. The apical cells of caulonemal filaments can grow 25-40 μm/h and divide every 6-8 hours (Schween *et al.*, 2003). *Physcomitrella patens* can be grown under axenic conditions on inorganic media. It has a very high capacity of regeneration. In routine culture, protonemal tissue can be disrupted and spread on solid or in liquid culture medium. The young protonemal tissue can be used to produce large numbers of protoplasts. Protoplasts can regenerate into protonemal tissues. In targeted gene knockout, polyethylene glycol (PEG)-mediated direct gene transfer into protoplasts is exploited (Schaefer *et al.*, 1991; Hohe *et al.*, 2004).

The genome size of *Physcomitrella* is about 480Mb, and it has 27 small chromosomes (Schaefer, 2002). There are about 120,000 ESTs are available. Comparative genomics study of the *Physcomitrella* gemetophytic transcriptome shows that at least 66% of *Arabidopsis thaliana* genes have homologues in *Physcomitrella* and more than 90% of the most closely related homologues of *Physcomitrella* gametophytic transcripts occur in vascular plants (Nishiyama *et al.*, 2003). Studies of plant-specific genes can be carried out in *Physcomitrella*. The genome sequence project of *Physcomitrella* has just been completed and been released to the public (http://genome.jgi-psf.org/Physcomitrella) (Quatrano *et al*., 2007). This will greatly facilitate the study in *Physcomitrella*.

Bryophytes are very old land plants. The moss is thought to have diverged from vascular plants immediately after green plants emerged on land. How far has this group diverged? A better understanding of cellulose biosynthesis may be achieved through this

study. Analysis of CesA genes can help us to understand the function of these genes, as well as the evolution of cellulose biosynthesis.

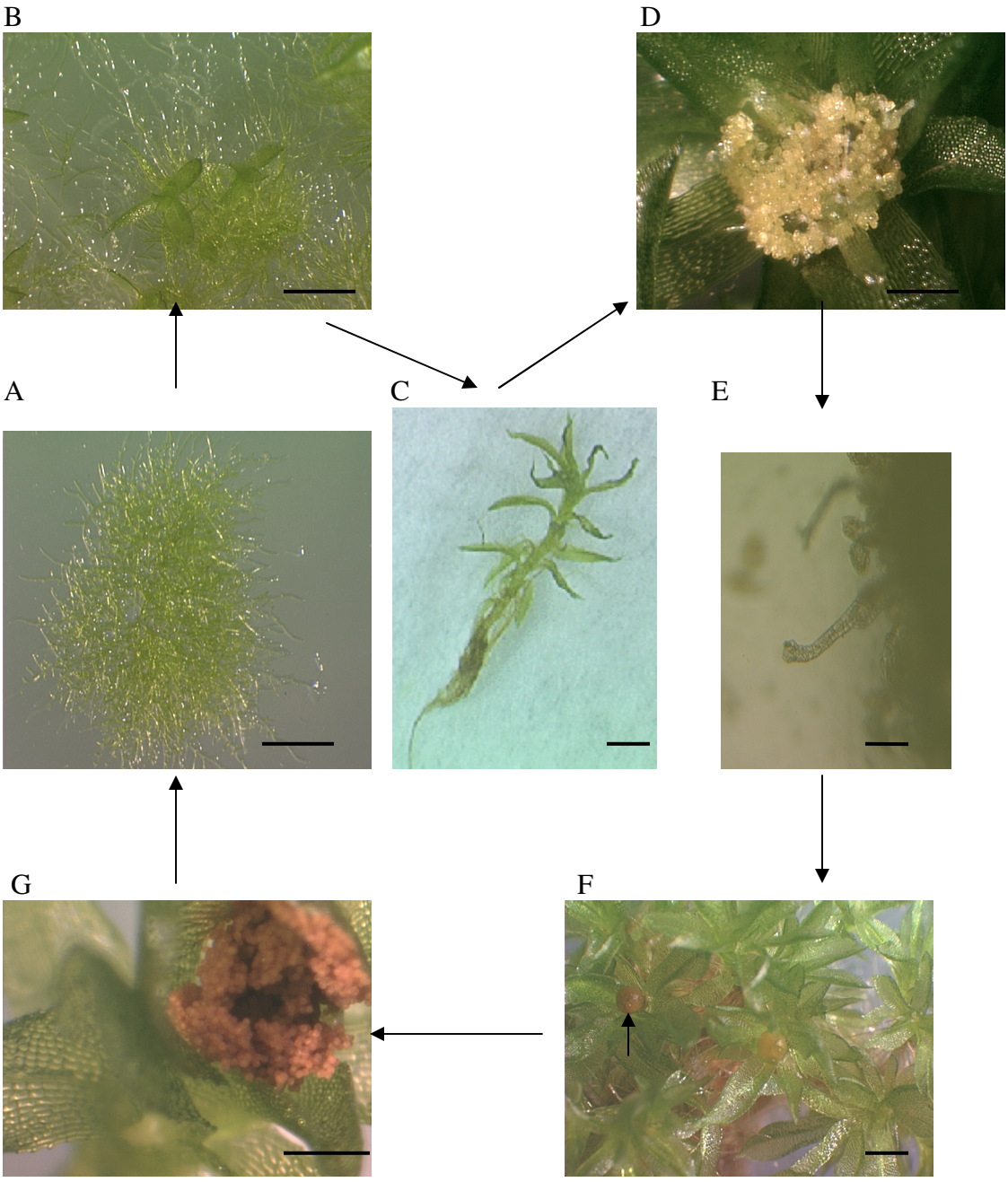


Figure 1-1 Life cycle of *Physcomitrella patens*

A. protonemata from germinated spore
B. protonemata with developing gametophore
C. gametophore
D. reproductive organs (a cluster of archegonia and antheridia
E. archegonia (single arrow) and antheridia (double arrows)
F. sporophyte
G. mature spores

Scale bar of A,B,C,D,F,G = 50mm Scale bar of E = 100μm

Figure 1-2 Cell walls of the chloronema and caulonema
The short arrow indicates a chloronema cell wall.
The long arrow indicates a caulonema cell wall.

Chapter 2 Identification and characterization of cellulose synthase genes in *Physcomitrella patens*

2.1 Introduction

Cellulose synthase (CesA) genes belong to a large gene family (Richmond and Somerville, 2000). Multigene family refers to a group of genes from the same organism that encode proteins with similar sequences either over their full lengths or limited to a specific domain. Multigene families are thought to evolve by gene duplication. Gene duplication has occurred repeatedly during the evolution of eukaryotes.

Mosses, liverworts and hornworts are called bryophytes. Together with ferns and seed plants, they are embryophytes (Parihar, 1961). Bryophytes are the only land plants with a dominant gametophyte (haploid) stage in their life history. Green plants emerged on land about 480 million years ago, and the mosses are thought to have diverged from vascular plants immediately after that time (Thiessen *et al*., 2001). Bryophytes were one of the earliest land plants and have several common characteristics with seed plants, such as very well-defined tissues, multicellular gametangia and an alternating life cycle. In classification of organisms, they have been placed between algae and true land plants (Smith, 1955).

There are 10 CesAs in *Arabidopsis*, at least 12 in maize, at least 10 in rice and at least 7 in aspen (Appenzeller *et al.*, 2004; Joshi *et al.*, 2004; Richmond, 2000; Tanaka *et al.*, 2003). How many cellulose synthase genes are present in *Physcomitrella*? How

similar are the *Physcomitrella* cellulose synthase(s) to higher plant cellulose synthases? My study of CesA genes in *Physcomitrella*, is designed to provide more knowledge about this gene family. It may help us understand more about how bryophytes and angiosperms relate to each other and how far these groups have diverged.

This chapter will discuss identification of cellulose synthase genes in *Physcomitrella patens*.

2.2 Materials and methods

Plant material

The strain of *Physcomitrella patens* used in the laboratories all over the world is the Gransden strain. It derives from a single spore collected by Harold Whitehouse at Gransden Wood Huntingdonshire, England in 1962 (Ashton and Cove, 1977). One plate of protonemata from germinated spores of this strain was obtained from the University of Washington. This plate of protonemata was further sub-cultured and was used as material for my research. For sub-culture, the protomemata were chopped by an autoclaved razor blade and then suspended in sterile water. About 1ml of suspension was spread on one Petri dish of BCDAT medium overlaid with cellophane. The plant material was cultured at 25^0C under 16hr light and 8 hr dark light cycle.

BCD and BCDAT media are routinely used in *Physcomitrella* culture. BCD medium contains 1mM $MgSO_4$, 10mM KNO_3, 45 μM $FeSO_4$, 1.8 mM KH_2PO_4, trace element solutions (0.22 μM $CuSO_4$, 0.19 μM $ZnSO_4$, 10 μM H_3BO_4, 0.10 μM Na_2MoO_4, 2 μM $MnCl_2$, 0.23 μM $CoCl_2$ and 0.17 μM KI), 1mM $CaCl_2$ [PH 6.5 adjusted with KOH).

0.8% (W/V) agar was added to this medium. BCDAT medium is the BCD medium supplemented with 5mM diammonium (+)-tartrate.

EST and genomic database search

PHYSCObase (http://moss.nibb.ac.jp) was the first *Physcomitrella* DNA database available to the public. This database contained EST sequences generated from the regenerating protoplasts library, the untreated protonemata library, the auxin-treated library, the cytokinin-treated library and the developing sporophyte library. *Arabidopsis* CesA sequences were used as queries to search against this database. Contigs with the e-value above 1.0 underwent further analysis. These contigs were submitted to NCBI (http://www.ncbi.nlm.nih.gov/) for BLASTX search. The sequences identified as CesAs were the most interesting ones. The EST clones of these sequences were obtained from RIKEN BioResource Center, Tsukuba-shi, Ibaraki, Japan and they were sequenced.

The entire genome shotgun project of *Physcomitrella patens* started from 2004 at the Joint Genome Institute of U.S. Department of Energy (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.info.html). All sequenced DNA data were released and deposited in GenBank during sequencing. Megablast was available to search against the the *Physcomitrella* genomic database. The identified CesA cDNA sequences were used in the megablast search (http://www.ncbi.nlm.nih.gov/blast).

All the searches were repeated after the release of more EST sequences and raw genomic sequences.

Genomic DNA sequences of PpCesA6 and PpCesA7

One week old sub-cultured protonemata growing on BCDAT medium were used for DNA extract. The DNA extract was carried by using a DNA DNeasy Plant Mini Kit (Qiagen; Valencia, CA, USA).

Primers gA6-5-1 5'-CAGCTCAATCGTCACGCGTCCTCCTGAGTCTTC-3'and gA6-3-1 5'-TACTTCCAGCACCGCACTCACATGCCCACATCC-3', derived from the 5'UTR and 3'UTR of PpCesA6 cDNA sequences respectively, were used to amplify the genomic DNA sequence of PpCesA6.

Primers gA7-5-1 5'-TCTTCTTCATGCCGTGGTGGTGAGCAGCAGGAG-3' and gA7-3-1 5-'ACCGCACTCACGTGCCCACATCCACCATTCAAC-3', derived from the 5'UTR and 3'UTR of PpCesA7 cDNA sequences respectively, were used to amplify the genomic DNA sequence of PpCesA7.

Herculase Hotstart DNA Polymerase (Stratagene; La Jolla, CA, USA) was used for PCR. The single-block temperature cycle was used. The PCR reaction was conducted in 50µl volume. The conditions for the reaction were: 5µl 10x reaction buffer, 1µl 10mM mixed dNTP, 2µl 12.5µm forward primer, 2µl 12.5µm reverse primer, 1.5µl genomic DNA (200ng), 1.5µl Herculase Hotstart DNA polymerase, 37µl H2O. For amplifying genomic DNA of PpCesA6, the annealing temperature was 61.5^0C. The reaction condition was 95^0C for 2 minutes, 10 cycles of 95^0C for 30s, 61.5^0C for 30s, 72^0C for 7 minutes, 25 cycles of 95^0C for 30s, 61.5^0C for 30s, 72^0C for 7 minutes plus 10s/cycle with a final extension at 72^0C for 10 minutes, followed by a 4^0C hold. For amplifying genomic DNA of PpCesA7, except that the annealing temperature was 64^0C, all the other

conditions were the same as for PpCesA6. After gel electrophoresis, the PCR products were purified with Qiaquick gel purification kit (Qiagen) and were sequenced on ABI 3730 DNA analyzer (DNA sequencing facility, the University of Texas at Austin).

DNA sequencing was conducted from both strands. Based on the sequenced part of the sequences, more primers were designed for further sequencing. Forward primers gA6-5-1, gA6-5-2, gA6-5-3, gA6-5-4, gA6-5-5, gA6-5-6, pdp45977-5-1, pdp45951-5-4, pdp45951-5-5-2, pdp45977-5-6, pdp45977-5-10, pdp45951-5-8, and reverse primers gA6-3-1, gA6-3-2, pdp45977-3-5, pdp45951-3-7, pdp45977-3-4 were used for sequencing PpCesA6 genomic DNA. Forward primers gA7-5-1, gA7-5-2, gA7-5-3, gA7-5-4, pdp45977-5-6, pdp45951-5-7, pdp45951-5-8, pdp45951-5-4, pdp45977-5-1, pdp45951-5-5-2 and reverse primers gA7-3-1, pdp45977-3-1, pdp45951-3-7, pdp45951-3-8, pdp45977-3-5, pdp45977-3-4, pdp45977-3-3 were used for sequencing PpCesA7 genomic DNA. The sequences of these primers are listed in table 2-2.

2.3 Results

2.3.1 PpCesA6 and PpCesA7 genes

(1) Gene structures

When this project first started in 2003, no expressed sequence tags (ESTs) and genomic databases were available for *Physcomitrella patens*. I obtained a cDNA library and a genomic DNA library from the University of Leeds (Quatrano *et al*., 1999). Degenerated primers derived from available CesA genes of *Arabidopsis* were used to amplify DNA fragments from these libraries. Several DNA fragments were obtained by

this method. Sequence analysis showed that they were partial CesA sequences. Then, in late 2003, about 80,000ESTs became available to the public through PHYSCObase (http://moss.nibb.ac.jp).

When using *Arabidopsis* CesAs as queries to search against the PHYSCObase, three putative transcripts gave the strongest hits. They matched the partial DNA sequences that were obtained from the cDNA library. Putative transcripts P010857 and P006926 were full-length cDNAs. They showed high similarity to each other and caused lots of interest. EST clones pphn50j06 and pphn19n05 (corresponding to the putative transcript P010857) and EST clones pphn50k09 and pphb2c24 (corresponding to the putative transcript P006926) were obtained from RIKEN BioResource Center, Japan (http://www.brc.riken.go.jp/lab/epd/catalog/p_patens.html) and underwent further analysis. Both strands of these clones were fully sequenced. The sequences obtained represented the sequences of the first two full-length cellulose synthase genes of *Physcomitrella* that had been identified. The full length sequences were named PpCesA6 and PpCesA7 and they were deposited in the GenBank (Accession number: DQ160224, DQ160225).

PpCesA6 cDNA is 4055 bp long with a coding region of 3288 bp (Figure2-1). The ORF starts from ATG at position 161 and stops at position 3451. The 5'-UTR contains 160 bp and the 3'-UTR contains 607 bp. PpCesA7 cDNA is 4000 bp long (Figure 2-2). The ORF starts from ATG at position 134 and stops at position 3424. Its coding region is also 3288 bp. The 5'-UTR contains 133 bp and the 3'-UTR contains 579 bp. The coding regions of PpCesA6 and PpCesA7 are almost 100% identical.

The identified full-length cDNA sequences were used to search against the shotgun genomic sequence database. I was unable to obtain the full genomic sequences of both genes due to incomplete sequencing of the genome (only 99% of the genome was sequenced) and the difficulty of assembling the genomic contigs, because of the high similarity of these two genes.

Based on the cDNA sequences of PpCesA6 and PpCesA7, primers were designed to amplify the genomic DNA sequences of both genes. The PCR product of genomic DNA sequence of PpCesA6 was larger than 8kb (Figure 2-3) and the PCR product of genomic DNA sequence of PpCesA7 was less than 8kb (Figure 2-4). Both PCR products were sequenced and the sequencing results were compared with the cDNA sequences. The resulting complete genomic sequences of PpCesA6 and PpCesA7 were deposited in the GenBank (Accession number: EF094116, EF094117).

The complete genomic DNA of PpCesA6 is 8160bp and PpCesA7 is 7739 bp. Both genes have 13 exons and 12 introns. Their first intron in both cases is more than 1kb long. All the other introns are small. The gene structures of PpCesA6 and PpCesA7 are shown in Figure 2-5. At the genomic DNA level, PpCesA6 and PpCesA7 also show high similarity. Except for the first intron and UTR regions, all the remaining regions of these two genes are almost identical.

(2) Protein structures

Both PpCesA6 and PpCesA7 encode for a protein with 1096 amino acids. The proteins encoded by PpCesA6 and PpCesA7 are different at only 3 positions (amino acid

numbers 8, 24 and 26). Nucleotide and derived amino acid sequences of full length cDNAs of PpCesA6 and PpCesA7 are shown in Figures 2-1 and 2-2, respectively. The conceptual translations of both sequences have 8 transmembrane helices predicted by HMMTOP (Tusnády and Simon, 1998, 2001).

PpCesA6 and PpCesA7 share 74% amino acid identity with Monterey pine (*Pinus radiate*) PrCesA10, 72% amino acid identity with quaking aspen (*Populus tremuloide*) PtrCesA6, 71% amino acid identity with maize (*Zea mays*) ZmCesA7 and three rice (*Oryza sativa*) CesAs, 65%-68% amino acid identity with *Arabidopsis* CesAs. The multiple alignments of the deduced amino acid sequences of PpCesA6 and PpCesA7 with several vascular plant CesA proteins by CLUSTALX (Thompson *et al.*, 1997) are shown in Figure 2-6. PpCesA6 and PpCesA7 contain the D, D, D, QXXRW motif in the form of DDG, DCD, TED, QVLRW, which is found in plant cellulose synthases. The Zinc-finger domain that contains reserved cysteine residues is found near the N-terminus in the form of $X_2CX_{15}CX_2CX_4CX_2CX_{11}CX_2C$ (from position38 to 83). This domain is suggested to be involved in protein-protein interaction between different CesA subunits (Kurek *et al.*, 2002).

2.3.2 CesA genes of *Physcomitrella patens*

More ESTs were available after PpCesA6 and PpCesA7 had been identified. Approximately 120,000 ESTs are accessible in the public domain now. When using the *Arabidopsis* CesAs as queries searching the database again, more putative transcripts were identified. Initially 15 putative transcripts were identified as CesAs. Further

analysis showed that 4 of them are CslD genes and some of the remaining putative transcripts belong to the same gene. 8 CesAs were identified (Table 2-1).

The EST clones of these putative transcripts were obtained from RIKEN BioResource Center. Most of these EST clones were sequenced. Another group of researchers were also working on identifying cellulose synthase genes. They identified 3 additional CesAs. All together, there are 11 CesAs in *Physcomitrella*. Partial or full nucleotide sequences of cellulose synthase genes were deposited in GenBank by this group (Roberts and Roberts, 2004; Roberts and Bushoven, 2007). In order to avoid repeated work, the sequenced result of CesAs (except CesA6 and CesA7) of my research was not deposited in Genbank. So far, mRNA or genomic DNA sequences of 10 of these genes were in the GenBank.

Like PpCesA6 and PpCesA7 that show high similarity, two other pairs of genes, PpCesA3 (GenBank accession number AY633543) and PpCesA8 (DQ902549), PpCesA4 (DQ902545) and PpCesA10 (DQ902551), also show high similarity. PpCesA2 (AY633540) and PpCesA9 (DQ902550) are pseudogenes.

2.4 Discussion

The cellulose synthase gene family has been studied intensively in vascular plants such as *Arabidopsis*, rice, maize and aspen (Holland *et al.*, 2000; Hazen *et al.*,2002; Joshi *et al.*, 2004; Richmond and Somerville, 2000; Tanaka *et al.*, 2003). Bryophytes are very old land plants and they are considered as missing links between algae and vascular plants (Smith, 1955). This study showed that there are 11 CesAs in *Physcomitrella*

patens. Except for two pseudogenes (PpCesA2 and PpCesA9), there are 9 functional CesAs. There are 10 cellulose synthase genes in *Arabidopsis* (Richmond, 2000), at least 12 in maize (Appenzeller *et al.*, 2004), at least 10 in rice (Tanaka *et al.*, 2003) and at least 7 in aspen (Joshi *et al.*, 2004). *Physcomitrella* has fewer cell types than angiosperms. Unlike vascular plants, which have both a primary and secondary cell wall, mosses only have a primary cell wall. Some bryophytes (especially mosses) have very well specialized cells, such as stereids, hydroids and leptoids, that are similar to vascular tissue. However, unlike vascular tissue, these cells lack lignin. Stereids have thickened cell walls, which are specialized in support. Hydroids are water-conducing cells and leptoids are sugar-conducting cells (Hebant, 1977). It is surprising that this moss has more than 10 cellulose synthase genes.

Among all the CesA genes in *Physcomitrella*, three pairs show high similarity (PpCesA6 and PpCesA7, PpCesA3 and PpCesA8, PpCesA4 and PpCesA10). This suggests that these genes may have arisen by recent gene duplication. Gene duplications of cellulose synthase genes are also observed in higher plants. In *Arabidopsis*, AtCesA-2 and -9 are very similar in sequence. In maize, ZmCesA-1 and -2, ZmCesA-4 and -9 are two pairs of very similar sequences (Holland *et al.*, 2000).

Gene duplication exists in all three domains of life: bacteria, archaebacteria and eukaryotes (Zhang, 2003). For instance, in *Arabidopsis*, there are more than 25,000 genes. About 65% of these genes, which accounts for more than 16,000 genes, are duplicated genes (The *Arabidopsis* genome initiative, 2000; Bowers *et al.*, 2003). Large gene duplication also occurred in rice (Paterson *et al.*, 2004). Analysis on the EST

sequences of *Physcomitrella* has shown that this moss is a paleopolyploid and the genome duplication may have occurred between 30 and 60 million years ago (Rensing *et al.*, 2007).

Gene duplication may occur by unequal crossing-over, retrotransposition or chromosomal duplication. The duplicated genes have many different fates. One of these is that the duplicated genes can maintain the same function. Gene duplication is a key mechanism in evolution. Once a gene is duplicated, the identical genes can undergo changes and diverge into two different genes. Usually the duplicated genes show expression divergence. Another fate of the duplicated genes is to become pseudogenes. If the gene dosage is not very critical, the duplicated gene can evolve to have a different function (Zhang, 2003).

Vascular plant cellulose synthase genes vary in size from 3.5kb to 5.5kb. They have 8 or 9 small introns (Richmond, 2000). PpCesA6 and PpCesA7 are large genes. They are more than 7kb long and they have 12 introns. The first intron of both genes is more than 1kb long. The deduced amino acid sequences of PpCesA6 and PpCesA7 share a maximum identity of 74% with vascular plant CesAs. Their coding regions are almost identical. They encode almost the same protein (they differ in only 3 amino acids). At the genomic level, except for the first introns, all the other regions of these two genes are almost identical.

Among the 10 *Aradidopsis* CesA genes, AtCesA1, 2, 3, 5, 6 are involved in primary cell wall synthesis. AtCesA1 and AtCesA3 are necessary while AtCesA2, AtCesA5 and AtCesA6 are partially redundant (Somerville, 2006; Desprez *et al.*, 2002,

2007; Persson *et al.*, 2007). PpCesA6 and PpCesA7 are almost identical; it is very likely that their functions are redundant, or at least partially redundant.

2.5 Acknowledgements

I would like to thank RIKEN Bioresource Center, Tsukuba-shi, Ibaraki, Japan for providing me EST clones. I would also like to thank Dr. Ralph Quatrano at the University of Washington for kindly providing me *Physcomitrella* protonemata.

```
                                                                a  1
gatgatcactcagaggtgaggtgcgaggaagaaaggggggacgagttcccgagtgttcttc 61
ccagtggcatgggtgctgggtgagaagaaggtggtttgtgagggagaacagggctcgagg 121
aggttggagccgtgggggcagtggtggagttagggaagc atg gaggccaatgcggggttg 181
                                        M  E  A  N  A  G  L  7
gtggcgggctcgcacaaccgcaatgagctggttgtcatccgacaggagagcgatgggccg 241
 V  A  G  S  H  N  R  N  E  L  V  V  I  R  Q  E  S  D  G  P  27
aggccgttgagtaatgtgaacagccacatctgtcaaatttgtggcgatgatgtgggagtg 301
 R  P  L  S  N  V  N  S  H  I  C  Q  I  C  G  D  D  V  G  V  47
acactggaaggggagatgttcgtggcgtgcaccgaatgcgggttcccggtgtgccgcccc 361
 T  L  E  G  E  M  F  V  A  C  T  E  C  G  F  P  V  C  R  P  67
tgctacgagtatgaacggaaggatgggactcaggcatgtccccagtgtcggactcgctac 421
 C  Y  E  Y  E  R  K  D  G  T  Q  A  C  P  Q  C  R  T  R  Y  87
aggcgacacaaagggagtcctcgagtgaagggagacgacgaagaagaggacaccgacgac 481
 R  R  H  K  G  S  P  R  V  K  G  D  D  E  E  E  D  T  D  D  107
ctggacaacgaattcaaccacaatgttgatatcgacaagcacgacaagcagcaagtcgtg 541
 L  D  N  E  F  N  H  N  V  D  I  D  K  H  D  K  Q  Q  V  V  127
gatgagatgctgcacagccagatggcgtatggtcgtgacacggacgtgatgatgtctgca 601
 D  E  M  L  H  S  Q  M  A  Y  G  R  D  T  D  V  M  M  S  A  147
atgcagcctcagtatccccttttgacggacggacacacggtttctggagcaggcgaatcg 661
 M  Q  P  Q  Y  P  L  L  T  D  G  H  T  V  S  G  A  G  E  S  167
aatgctacgtcacccgatcatcaggcgatatttcctgttgcaggagggaaacggatccac 721
 N  A  T  S  P  D  H  Q  A  I  F  P  V  A  G  G  K  R  I  H  187
cctgttgcttacagcgacattggaagcccagctaggccactggatccggcgaaggatctg 781
 P  V  A  Y  S  D  I  G  S  P  A  R  P  L  D  P  A  K  D  L  207
ggatcgtatgggtacgggagcatcgcctggaaggagagggtggagagctggaagctgagg 841
 G  S  Y  G  Y  G  S  I  A  W  K  E  R  V  E  S  W  K  L  R  227
cagggaatgcaaatgacgacgacggagggagggcagctccaagcaagcgggaagggcggc 901
 Q  G  M  Q  M  T  T  T  E  G  G  Q  L  Q  A  S  G  K  G  G  247
cacgatgagaacgggccagattgcccggacctgccaataatggacgaatcgcgacaacca 961
 H  D  E  N  G  P  D  C  P  D  L  P  I  M  D  E  S  R  Q  P  267
ctgtcgaggaaagtgccgattccatcgagcaagatcaacccgtacaggatgataattgtg 1021
 L  S  R  K  V  P  I  P  S  S  K  I  N  P  Y  R  M  I  I  V  287
attcgactggtggtgatttgtctgtttttccgataccgtatcttgaaccctgtgaatgaa 1081
 I  R  L  V  V  I  C  L  F  F  R  Y  R  I  L  N  P  V  N  E  307
gcgtacgccctgtggctggtgtcggtgatttgcgagatttggttcgccatatcgtggatt 1141
 A  Y  A  L  W  L  V  S  V  I  C  E  I  W  F  A  I  S  W  I  327
ctggatcagtttccgaagtggctgcccatcaaccgcgagacgtatctggatcggctgtcg 1201
 L  D  Q  F  P  K  W  L  P  I  N  R  E  T  Y  L  D  R  L  S  347
ttgaggttcgagaaggaaggggagccgtctcggctgtgtcccgtggatatctatgtgagt 1261
 L  R  F  E  K  E  G  E  P  S  R  L  C  P  V  D  I  Y  V  S  367
acggtggacccaatgaaggagccacctctggtgactgcgaacacgatactgtcgattctg 1321
 T  V  D  P  M  K  E  P  P  L  V  T  A  N  T  I  L  S  I  L  387
gctgtggactaccccgtggacaaggtgtcgtgctacatatccgatgatggtgcatcgatg 1381
 A  V  D  Y  P  V  D  K  V  S  C  Y  I  S  D  D  G  A  S  M  407
ctgacgttcgaggttctgtcggagacgtcggagttcgctcggaagtgggtgccgttctgc 1441
 L  T  F  E  V  L  S  E  T  S  E  F  A  R  K  W  V  P  F  C  427
aagaagttcaacatcgagccccgggcacctgaggtgtacttcgccctgaagattgactac 1501
 K  K  F  N  I  E  P  R  A  P  E  V  Y  F  A  L  K  I  D  Y  447
ctgaaggacaaggtgcagccaacgttcgtgaaggagcggagagccatgaagagggagtac 1561
 L  K  D  K  V  Q  P  T  F  V  K  E  R  R  A  M  K  R  E  Y  467
```

```
gaggagttcaaggtgcgagtgaacgcgttggtggcgaaagcgcagaagatgccggacgaa 1621
 E  E  F  K  V  R  V  N  A  L  V  A  K  A  Q  K  M  P  D  E  487
ggatggacaatgcaggatggaacgccgtggcctgggaacaacactcgcgaccatcccggg 1681
 G  W  T  M  Q  D  G  T  P  W  P  G  N  N  T  R  D  H  P  G  507
atgatccaggtgtttttgggacacagcgggggtcacgacacggacgggaacgagctgcca 1741
 M  I  Q  V  F  L  G  H  S  G  G  H  D  T  D  G  N  E  L  P  527
cggctggtgtacgtgtctcgagagaaacgacccgggttcaaccatcacaagaaggccggt 1801
 R  L  V  Y  V  S  R  E  K  R  P  G  F  N  H  H  K  K  A  G  547
gccatgaatgcgttggtgcgggtgtctgcggtgctaacaaacgcgccttttcttcctgaat 1861
 A  M  N  A  L  V  R  V  S  A  V  L  T  N  A  P  F  F  L  N  567
ctggattgtgatcattacatcaacaacagcaaggcgctccgggaagcgatgtgctttctg 1921
 L  D  C  D  H  Y  I  N  N  S  K  A  L  R  E  A  M  C  F  L  587
atggatcccatcgtggggaagagggtgtgctacgtccagtttcctcagcggttcgatggc 1981
 M  D  P  I  V  G  K  R  V  C  Y  V  Q  F  P  Q  R  F  D  G  607
atcgacaggaacgatcgatatgccaatcacaacaccgttttcttcgacatcaacttgaag 2041
 I  D  R  N  D  R  Y  A  N  H  N  T  V  F  F  D  I  N  L  K  627
ggattggacggcgtgcagggcccggtgtacgtgggtactggatgttgtttcaagaggcaa 2101
 G  L  D  G  V  Q  G  P  V  Y  V  G  T  G  C  C  F  K  R  Q  647
gcgatttatggttacgaccctcctccgaaggatgcgaaggcgtcgggtgggcggagccaa 2161
 A  I  Y  G  Y  D  P  P  P  K  D  A  K  A  S  G  G  R  S  Q  667
ggcgtgtgtccatcatggctgtgcgggccccggaagaagggagttgggaaggcgaaggtt 2221
 G  V  C  P  S  W  L  C  G  P  R  K  K  G  V  G  K  A  K  V  687
gcgaaaggcgggaagaagaagcctccgtcgaggagcgactccagcattcccattttcagc 2281
 A  K  G  G  K  K  K  P  P  S  R  S  D  S  S  I  P  I  F  S  707
ctggaggacatcgaagaaggcatcgaaggcattgacgaggagaagtcgtcgctgatgtct 2341
 L  E  D  I  E  E  G  I  E  G  I  D  E  E  K  S  S  L  M  S  727
ttgaagaacttcgagaagaggttcggtcagtctccggtgttcgtggcgtcgacgctgctg 2401
 L  K  N  F  E  K  R  F  G  Q  S  P  V  F  V  A  S  T  L  L  747
gagaacggaggcgtgccgcactctgcgaatccggggtcgctgttgaaggaagccatccac 2461
 E  N  G  G  V  P  H  S  A  N  P  G  S  L  L  K  E  A  I  H  767
gtgatcagttgtgggtacgaagacaagacggactgggggaaggagatcggatggatctac 2521
 V  I  S  C  G  Y  E  D  K  T  D  W  G  K  E  I  G  W  I  Y  787
gggtccgtgacggaggacattctgacggggttcaaaatgcactgcagaggatggaggtcc 2581
 G  S  V  T  E  D  I  L  T  G  F  K  M  H  C  R  G  W  R  S  807
atctattgcatgccgacgcgacctgcgttcaaggggtcggcgcccatcaacttgtcggat 2641
 I  Y  C  M  P  T  R  P  A  F  K  G  S  A  P  I  N  L  S  D  827
cggctgaaccaagtgctgcgatgggcgctggggtcggtggagatttcgctgagtcggcac 2701
 R  L  N  Q  V  L  R  W  A  L  G  S  V  E  I  S  L  S  R  H  847
tgccctctgtggtacgggtacggtggagggaagaacggcgggctgaagtgtctggagagg 2761
 C  P  L  W  Y  G  Y  G  G  G  K  N  G  G  L  K  C  L  E  R  867
ctggcttacatcaacacgacgatctacccgctgacatcgttgccgctgctggcgtactgc 2821
 L  A  Y  I  N  T  T  I  Y  P  L  T  S  L  P  L  L  A  Y  C  887
gtgctgcccgcggtgtgtttgctgacggggaagttcatcatacccacgatcagtaacctg 2881
 V  L  P  A  V  C  L  L  T  G  K  F  I  I  P  T  I  S  N  L  907
gcgagtctgtggtttatttcattgtttatctccatttttgcgacgggcatactggaaatg 2941
 A  S  L  W  F  I  S  L  F  I  S  I  F  A  T  G  I  L  E  M  927
cggtggtcgggagtgggcatcgacgagtggtggaggaacgagcagttttgggtgatcgga 3001
 R  W  S  G  V  G  I  D  E  W  W  R  N  E  Q  F  W  V  I  G  947
ggcgtgtcggcgcacttgttcgcgctgttccagggtctgctgaaggtgttcgcgggtatc 3061
 G  V  S  A  H  L  F  A  L  F  Q  G  L  L  K  V  F  A  G  I  967
gacaccaacttcacggtgacgtcgaaacaggcagaggacgaggatttcgcggagctgtac 3121
 D  T  N  F  T  V  T  S  K  Q  A  E  D  E  D  F  A  E  L  Y  987
atgataaagtggacggcgctgctgattccgccgacgacgctgatcgtgatcaacatgatc 3181
 M  I  K  W  T  A  L  L  I  P  P  T  T  L  I  V  I  N  M  I  1007
```

```
ggcgtggtggcggggatttcggatgcgatcaacaacggataccagtcgtggggcccgctg 3241
 G  V  V  A  G  I  S  D  A  I  N  N  G  Y  Q  S  W  G  P  L  1027
ttcgggaagctgtttttcgcgttctgggtgatcgttcacttgtatccgttcctgaagggt 3301
 F  G  K  L  F  F  A  F  W  V  I  V  H  L  Y  P  F  L  K  G  1047
ctgatgggacgacagaaccggacaccgacgatcgtgatcgtgtggtcgatcctgctggcg 3361
 L  M  G  R  Q  N  R  T  P  T  I  V  I  V  W  S  I  L  L  A  1067
tccatcttctcgcttctgtgggtgcgcatcgacccgtttctggcgaaggtgaagggtccc 3421
 S  I  F  S  L  L  W  V  R  I  D  P  F  L  A  K  V  K  G  P  1087
gacctgtcgcagtgcgggataaactgttgagtgggggtgtgtgatggaggaatcgaggaa 3481
 D  L  S  Q  C  G  I  N  C  -                                 1096
atggatgttaggagtgcgacgctttggaacgcaacgtcaaggttctttggattgcgacgc 3541
ttgggcggggcgatgattgttttccgtggtgcacggaggaagttggttaaattataaggt 3601
tggcggatcacggccgaagaagcctagatgcgcgctctggcgcggagtcgtgaatatgaa 3661
gcaaggttccgcattggcggtatcgaggtgggcgacgaagaggtttgtagataataagct 3721
cgagtcgaggttgggtagatacaggttggggagggtttgacgagagatggactaacaagt 3781
aggtgggtgagtctagagtaggtagtagcacggtggattgtgttggtcgattggcaagcg 3841
tttctgttatccatattgccgttggagactttggtagacgtgtgtcggggtttgtgtctc 3901
ggtgcagctccaaatgaaactggtgccctggttgggtgggaagggtagcagaggggtttg 3961
ctcatacaaattaggccgttgcgtgaagcaggtgccatattgtgcagtatgccggttgtg 4021
aatggaaagatatcgtgaatcggatgttgaatggaaaaaaaaaaaaaaa             4070
```

Figure 2-1 Nucleotide and amino acid sequences of full length cDNA of PpCesA6.
The bold and underlined denote the conserved regions (D, DXD, D, and QXXRW). The reading frame starts at position 161(bold) and stops at position 3451(-).
The Zinc-finger domain that contains reserved cysteine residues is found near the N-terminus in the form of $X_2CX_{15}CX_2CX_4CX_2CX_{11}CX_2C$ (from position38 to 83).
The D, D, D, QXXRW motif is in the form of DDG (position 402), DCD (position 569), TED (position 791), QVLRW (position 831).

```
                                                                  g 1
acagatgatcactctcgggtgagaagggagtgcgcgggacggagttcccatgggggagag 61
tggcttgtgaggaagtgcgggaactaatcacgtgaggagaggctgcattgtggtggaagg 121
tggtttggcggc atg gaggcgaatgcagggctgctggcgggctcgcacaaccggaacgag 181
             M  E  A  N  A  G  L  L  A  G  S  H  N  R  N  E  16
ctggttgttattcggcaggagggtgatgagccgaggccgttgagtaatgtgaacagccac 241
 L  V  V  I  R  Q  E  G  D  E  P  R  P  L  S  N  V  N  S  H  36
atctgtcaaatttgtggcgacgatgtgggagtgacactggaaggggagatgttcgtggcg 301
 I  C  Q  I  C  G  D  D  V  G  V  T  L  E  G  E  M  F  V  A  56
tgcaccgaatgcgggttcccggtgtgccgcccctgctacgagtatgaacggaaggatggg 361
 C  T  E  C  G  F  P  V  C  R  P  C  Y  E  Y  E  R  K  D  G  76
actcaggcatgtccccagtgtcggactcgctacaggcgacacaaagggagtcctcgagtg 421
 T  Q  A  C  P  Q  C  R  T  R  Y  R  R  H  K  G  S  P  R  V  96
aagggagacgacgaagaagaggacaccgacgacctggacaacgaattcaaccacaatgtt 481
 K  G  D  D  E  E  E  D  T  D  D  L  D  N  E  F  N  H  N  V  116
gatatcgacaagcacgacaagcagcaagtcgtggatgagatgctgcacagccagatggcg 541
 D  I  D  K  H  D  K  Q  Q  V  V  D  E  M  L  H  S  Q  M  A  136
tatggtcgtgacacggacgtgatgatgtctgcaatgcagcctcagtatccccttttgacg 601
 Y  G  R  D  T  D  V  M  M  S  A  M  Q  P  Q  Y  P  L  L  T  156
gacggacacacggtttctggagcaggcgaatcgaatgctacgtcacccgatcatcaggcg 661
 D  G  H  T  V  S  G  A  G  E  S  N  A  T  S  P  D  H  Q  A  176
atatttcctgttgcaggagggaaacggatccaccctgttgcttacagcgacattggaagc 721
 I  F  P  V  A  G  G  K  R  I  H  P  V  A  Y  S  D  I  G  S  196
ccagctaggccactggatccggcgaaggatctgggatcgtatgggtacgggagcatcgcc 781
 P  A  R  P  L  D  P  A  K  D  L  G  S  Y  G  Y  G  S  I  A  216
tggaaggagagggtggagagctggaagctgaggcagggaatgcaaatgacgacgacggag 841
 W  K  E  R  V  E  S  W  K  L  R  Q  G  M  Q  M  T  T  T  E  236
ggagggcagctccaagcaagcgggaagggcggccacgatgagaacgggccagattgcccg 901
 G  G  Q  L  Q  A  S  G  K  G  G  H  D  E  N  G  P  D  C  P  256
gacctgccaataatggacgaatcgcgacaaccactgtcgaggaaagtgccgattccatcg 961
 D  L  P  I  M  D  E  S  R  Q  P  L  S  R  K  V  P  I  P  S  276
agcaagatcaacccgtacaggatgataattgtgattcgactggtggtgatttgtctgttt 1021
 S  K  I  N  P  Y  R  M  I  I  V  I  R  L  V  V  I  C  L  F  296
ttccgataccgtatcttgaaccctgtgaatgaagcgtacgccctgtggctggtgtcggtg 1081
 F  R  Y  R  I  L  N  P  V  N  E  A  Y  A  L  W  L  V  S  V  316
atttgcgagatttggttcgccatatcgtggattctggatcagtttccgaagtggctgccc 1141
 I  C  E  I  W  F  A  I  S  W  I  L  D  Q  F  P  K  W  L  P  336
atcaaccgcgagacgtatctggatcggctgtcgttgaggttcgagaaggaaggggagccg 1201
 I  N  R  E  T  Y  L  D  R  L  S  L  R  F  E  K  E  G  E  P  356
tctcggctgtgtcccgtggatatctatgtgagtacggtggacccaatgaaggagccacct 1261
 S  R  L  C  P  V  D  I  Y  V  S  T  V  D  P  M  K  E  P  P  376
ctggtgactgcgaacacgatactgtcgattctggctgtggactaccccgtggacaaggtg 1321
 L  V  T  A  N  T  I  L  S  I  L  A  V  D  Y  P  V  D  K  V  396
tcgtgctacatatccgatgatggtgcatcgatgctgacgttcgaggttctgtcggagacg 1381
 S  C  Y  I  S  D  D  G  A  S  M  L  T  F  E  V  L  S  E  T  416
tcggagttcgctcggaagtgggtgccgttctgcaagaagttcaacatcgagccccgggca 1441
 S  E  F  A  R  K  W  V  P  F  C  K  K  F  N  I  E  P  R  A  436
cctgaggtgtacttcgccctgaagattgactacctgaaggacaaggtgcagccaacgttc 1501
 P  E  V  Y  F  A  L  K  I  D  Y  L  K  D  K  V  Q  P  T  F  456
gtgaaggagcggagagccatgaagagggagtacgaggagttcaaggtgcgagtgaacgcg 1561
 V  K  E  R  R  A  M  K  R  E  Y  E  E  F  K  V  R  V  N  A  476
ttggtggcgaaagcgcagaagatgccggacgaaggatggacaatgcaggatggaacgccg 1621
 L  V  A  K  A  Q  K  M  P  D  E  G  W  T  M  Q  D  G  T  P  496
```

```
tggcctgggaacaacactcgcgaccatcccgggatgatccaggtgtttttgggacacagc 1681
 W  P  G  N  N  T  R  D  H  P  G  M  I  Q  V  F  L  G  H  S  516
gggggtcacgacacggacgggaacgagctgccacggctggtgtacgtgtctcgagagaaa 1741
 G  G  H  D  T  D  G  N  E  L  P  R  L  V  Y  V  S  R  E  K  536
cgacccgggttcaaccatcacaagaaggccggtgccatgaatgcgttggtgcgggtgtct 1801
 R  P  G  F  N  H  H  K  K  A  G  A  M  N  A  L  V  R  V  S  556
gcggtgctaacaaacgcgcctttcttcctgaatctggattgtgatcattacatcaacaac 1861
 A  V  L  T  N  A  P  F  F  L  N  L  D  C  D  H  Y  I  N  N  576
agcaaggcgctccgggaagcgatgtgctttctgatggatcccatcgtggggaagagggtg 1921
 S  K  A  L  R  E  A  M  C  F  L  M  D  P  I  V  G  K  R  V  596
tgctacgtccagtttcctcagcggttcgatggcatcgacaggaacgatcgatatgccaat 1981
 C  Y  V  Q  F  P  Q  R  F  D  G  I  D  R  N  D  R  Y  A  N  616
cacaacaccgttttcttcgacatcaacttgaagggattggacggcgtgcagggcccggtg 2041
 H  N  T  V  F  F  D  I  N  L  K  G  L  D  G  V  Q  G  P  V  636
tacgtgggtactggatgttgtttcaagaggcaagcgatttatggttacgaccctcctccg 2101
 Y  V  G  T  G  C  C  F  K  R  Q  A  I  Y  G  Y  D  P  P  P  656
aaggatgcgaaggcgtcgggtgggcggagccaaggcgtgtgtccatcatggctgtgcggg 2161
 K  D  A  K  A  S  G  G  R  S  Q  G  V  C  P  S  W  L  C  G  676
ccccggaagaagggagttgggaaggcgaaggttgcgaaaggcgggaagaagaagcctccg 2221
 P  R  K  K  G  V  G  K  A  K  V  A  K  G  G  K  K  K  P  P  696
tcgaggagcgactccagcattcccattttcagcctggaggacatcgaagaaggcatcgaa 2281
 S  R  S  D  S  S  I  P  I  F  S  L  E  D  I  E  E  G  I  E  716
ggcattgacgaggagaagtcgtcgctgatgtctttgaagaacttcgagaagaggttcggt 2341
 G  I  D  E  E  K  S  S  L  M  S  L  K  N  F  E  K  R  F  G  736
cagtctccggtgttcgtggcgtcgacgctgctggagaacggaggcgtgccgcactctgcg 2401
 Q  S  P  V  F  V  A  S  T  L  L  E  N  G  G  V  P  H  S  A  756
aatccggggtcgctgttgaaggaagccatccacgtgatcagttgtgggtacgaagacaag 2461
 N  P  G  S  L  L  K  E  A  I  H  V  I  S  C  G  Y  E  D  K  776
acggactgggggaaggagatcggatggatctacgggtccgtgacggaggacattctgacg 2521
 T  D  W  G  K  E  I  G  W  I  Y  G  S  V  T  E  D  I  L  T  796
gggttcaaaatgcactgcagaggatggaggtccatctattgcatgccgacgcgacctgcg 2581
 G  F  K  M  H  C  R  G  W  R  S  I  Y  C  M  P  T  R  P  A  816
ttcaaggggtcggcgcccatcaacttgtcggatcggctgaaccaagtgctgcgatgggcg 2641
 F  K  G  S  A  P  I  N  L  S  D  R  L  N  Q  V  L  R  W  A  836
ctggggtcggtggagatttcgctgagtcggcactgccctctgtggtacgggtacggtgga 2701
 L  G  S  V  E  I  S  L  S  R  H  C  P  L  W  Y  G  Y  G  G  856
gggaagaacggcgggctgaagtgtctggagaggctggcttacatcaacacgacgatctac 2761
 G  K  N  G  G  L  K  C  L  E  R  L  A  Y  I  N  T  T  I  Y  876
ccgctgacatcgttgccgctgctggcgtactgcgtgctgcccgcggtgtgtttgctgacg 2821
 P  L  T  S  L  P  L  L  A  Y  C  V  L  P  A  V  C  L  L  T  896
gggaagttcatcatacccacgatcagtaacctggcgagtctgtggtttatttcattgttt 2881
 G  K  F  I  I  P  T  I  S  N  L  A  S  L  W  F  I  S  L  F  916
atctccatttttgcgacgggcatactggaaatgcggtggtcgggagtgggcatcgacgag 2941
 I  S  I  F  A  T  G  I  L  E  M  R  W  S  G  V  G  I  D  E  936
tggtggaggaacgagcagttttgggtgatcggaggcgtgtcggcgcacttgttcgcgctg 3001
 W  W  R  N  E  Q  F  W  V  I  G  G  V  S  A  H  L  F  A  L  956
ttccagggtctgctgaaggtgttcgcgggtatcgacaccaacttcacggtgacgtcgaaa 3061
 F  Q  G  L  L  K  V  F  A  G  I  D  T  N  F  T  V  T  S  K  976
caggcagaggacgaggatttcgcggagctgtacatgataaagtggacggcgctgctgatt 3121
 Q  A  E  D  E  D  F  A  E  L  Y  M  I  K  W  T  A  L  L  I  996
ccgccgacgacgctgatcgtgatcaacatgatcggcgtggtggcggggatttcggatgcg 3181
 P  P  T  T  L  I  V  I  N  M  I  G  V  V  A  G  I  S  D  A 1016
atcaacaacggataccagtcgtggggcccgctgttcgggaagctgtttttcgcgttctgg 3241
 I  N  N  G  Y  Q  S  W  G  P  L  F  G  K  L  F  F  A  F  W 1036
```

```
gtgatcgttcacttgtatccgttcctgaagggtctgatgggacgacagaaccggacaccg 3301
 V  I  V  H  L  Y  P  F  L  K  G  L  M  G  R  Q  N  R  T  P  1056
acgatcgtgatcgtgtggtcgatcctgctggcgtccatcttctcgcttctgtgggtgcgc 3361
 T  I  V  I  V  W  S  I  L  L  A  S  I  F  S  L  L  W  V  R  1076
atcgacccgtttctggcgaaggtgaagggtcccgacctgtcgcagtgcgggataaactgt 3421
 I  D  P  F  L  A  K  V  K  G  P  D  L  S  Q  C  G  I  N  C  1096
tgagtgggggtgtgtgatggaggaatcgaggaaatggatgttaggagtgcgacgctttgg 3481
 –
aacgcaacgtcaaggttctttggattgcgacgcttgggcggggcgatgattgttttccgt 3541
ggtgcacggaggaagttggttaaattatgaggttggcgtatcacggcggaagaaacccag 3601
atgcgcatttgttaacggtggtgttgtactggaggaggtccccgcatgggcgacgaattg 3661
gtttgtagataataaaggagagataaggtgtgtcagatacttgcgaggagtcctatacga 3721
gagaagaccaatagttataagggtgagtgtagagtaggtagtagcacggtggattgtgtt 3781
ggttgcgtgccacacgatgttgttttccaggttgtcgttgtagcctcctttaggcagatg 3841
tcgggctttgtgtctcggtgcagccacagtttaaaccggtgccctggtagggtgggagga 3901
gtggcaggggggttcactcatacaaattaggccgtggcgtgaagcaagtgccatattgtg 3961
cagtatgccggttgtgaatggaaagatatcgtgaatcggaaaaaaaaaaaaaaaa      4016
```

Figure 2-2 Nucleotide and amino acid sequences of full length cDNA of PpCesA7.
The reading frame starts at position 134(bold) and stops at position 3424(-). The bold and underlined denote the conserved regions (D, DXD, D, and QXXRW).
The Zinc-finger domain that contains reserved cysteine residues is found near the N-terminus in the form of $X_2CX_{15}CX_2CX_4CX_2CX_{11}CX_2C$ (from position38 to 83).
The D, D, D, QXXRW motif is in the form of DDG (position 402), DCD (position 569), TED (position 791), QVLRW (position 831).

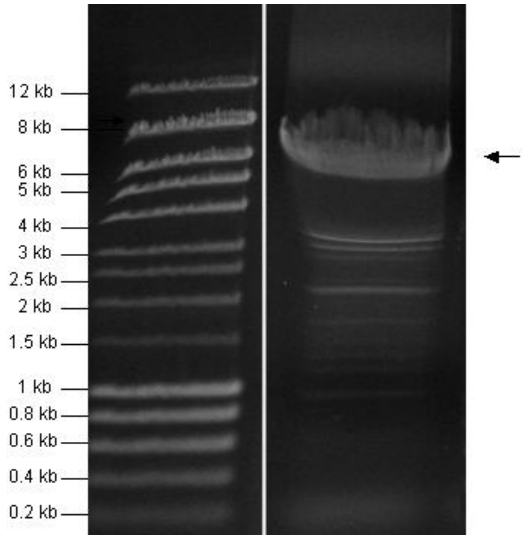


Figure 2-3 PCR product of the amplified PpCesA6 genomic fragment
Arrow indicates the amplified fragment of 8.1 kb

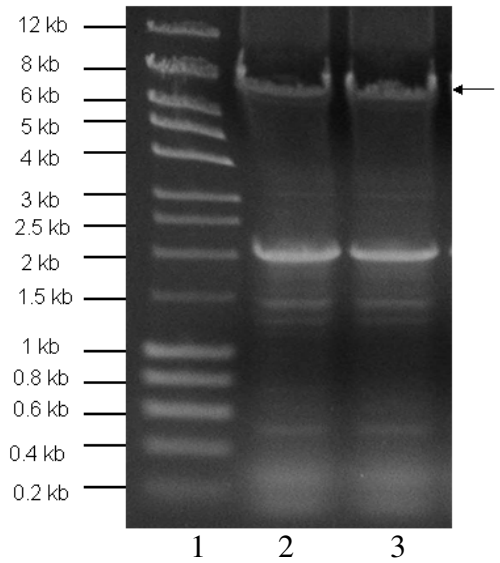


Figure 2-4 PCR product of the amplified PpCesA7 genomic fragment
Arrow indicates the amplified fragment of 7.7 kb
Lane 1, DNA marker.
Lane 2 and Lane 3, the amplified genomic fragment

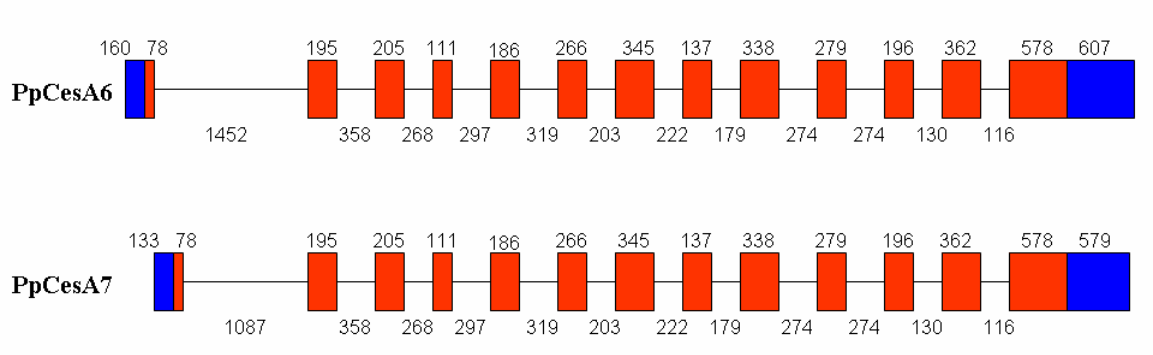
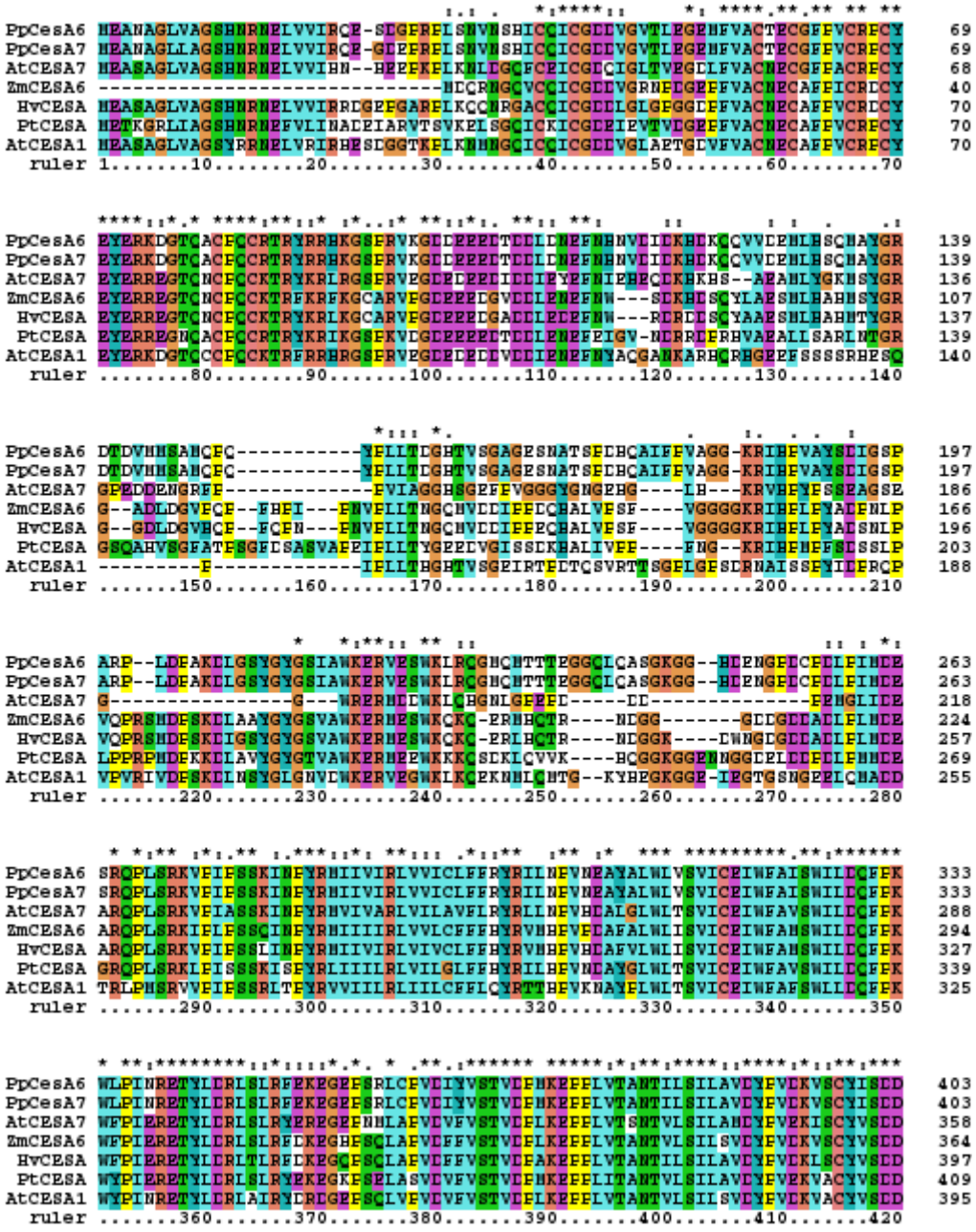
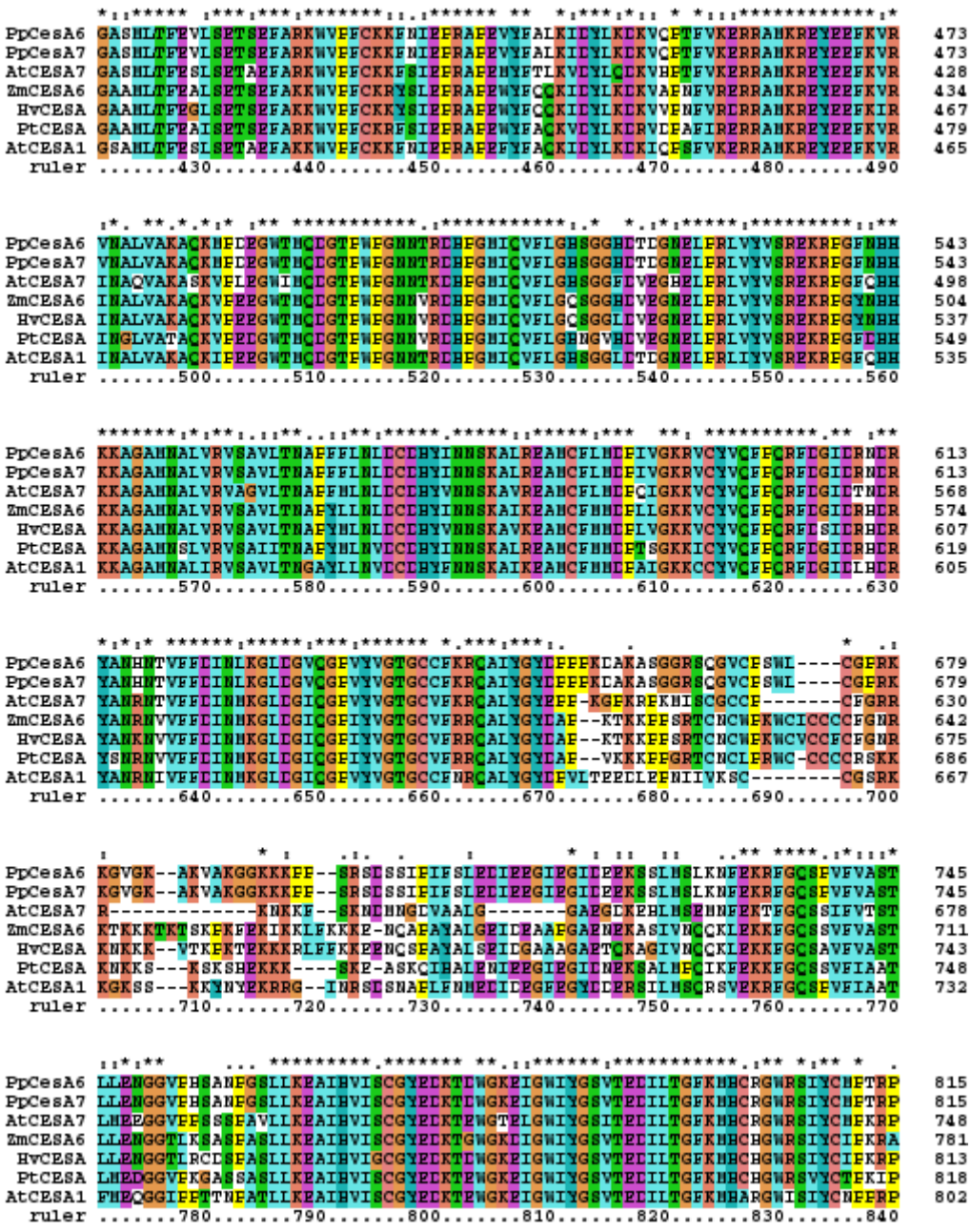


Figure 2-5 Gene Structures of PpCesA6 and PpCesA7
Red rectangles represent exons.
Blue rectangles represent UTRs.
Lines represent introns.
Note that except for the UTRs and the first introns, all the other regions of both genes are identical.





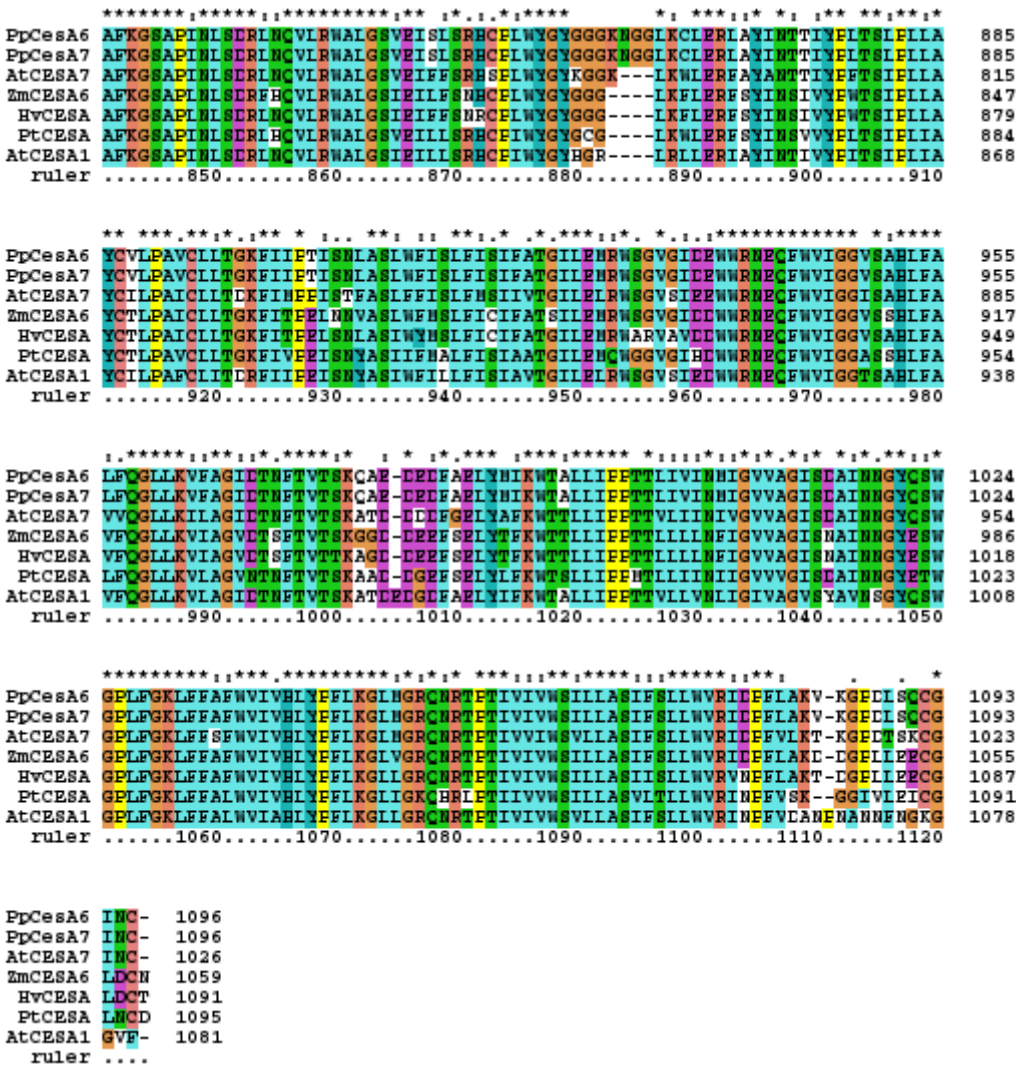


Figure 2-6 Multiple alignments of the deduced amino acids among PpCesA6, PpCesA7, some of the known plant cellulose synthases

PpCesA6, PpCesA7= *Physcomitrella patens* CesA6, CesA7 (GenBank accession number: AAZ86086, AAZ86087)
AtCesA1, AtCesA7= *Arabidopsis thaliana* CesA1, CesA7 (GenBank accession number: NP_194967, NP_197244)
HvCesA= *Hordeum vulgare* CesA (GenBank accession number AAR29964)
PtCesA= *Populus tremula x Populus tremuloides* CesA (GenBank accession number AAT09895)
ZmCesA6=*Zea mays* CesA6 (GenBank accession number AAF89966)

Note: All conserved residues (D, D, D) and motif (QXXRW) were found in all species shown here.

Table 2-1 *Physcomitrella patens* cellulose synthase genes

Gene	Putative transcript	EST clone	BRC stock number
PpCesA4	P025110	pphb20c07	pdp17615
PpCesA5	P024669	pphb1j16	pdp10761
	P016430	pphb45n02	pdp27094
		pphb40c17	pdp24996
PpCesA6	P010857	pphb19n05	pdp17495
		pphn50j06	pdp45951
PpCesA7	P006926	pphn50k09	pdp45977
		pphb2c24	pdp10967
PpCesA8	P011363	pphn29k01	pdp39044
		pphn33i03	pdp40456
PpCesA9	P013349	pphb45k12	pdp27034
		pphb17p17	pdp16811
		pphb17p18	pdp16812
PpCesA10	P016428	pphb45k22	pdp27044
PpCesA11	M.S.seq.1		

Putative transcripts were obtained from the PHYSCObase (http://moss.nibb.ac.jp). EST clones were obtained from BRC-RIKEN BioResources Center (http://www.brc.riken.go.jp/lab/epd/catalog/p_patens.html).

Table 2-2 DNA Sequencing Primers

gA6-5-1	5’-CAGCTCAATCGTCACGCGTCCTCCTGAGTCTTC- 3’
gA6-5-2	5’-CAGACTTGATCTGTAGGTTCT-3’
gA6-5-3	5’-GAAGTGGACACCAGTCTC-3’
gA6-5-4	5’-CTACCTGAAGGACAAGGTGC-3’
gA6-5-5	5’-AATGCAGGATGGAACGCCGTG-3’
gA6-5-6	5’-GCGAATCGAATGCTACGTCAC-3’
gA7-5-1	5’-TCTTCTTCATGCCGTGGTGGTGAGCAGCAGGAG- 3’
gA7-5-2	5’-GTTGGCCGTGCGTGTCATTCG-3’
gA7-5-3	5’-CAACTGAGTCTGTCACTCCTC-3’
gA7-5-4	5’-GTGGAAGTGGAACACCAGTCT-3’
pdp45951-5-4	5’-GTGGAGAGCTGGAAGCTGAGGCA-3’
pdp45951-5-5-2	5’-TGGCTGGTGTCGGTGATTTG-3’
pdp45951-5-7	5’-ATATCTATGTGAGTACGGTGGAC-3’
pdp45951-5-8	5’-AAGAAGGCCGGTGCCATGAATGC-3’
pdp45977-5-1	5’-ATGCTGCACAGCCAGATGGCGTA-3’
pdp45977-5-6	5’-CGTGTGTCCATCATGGCTGTGCG-3’
pdp45977-5-10	5’-ACCAAGTGCTGCGATGGGCGCTG-3’
gA6-3-1	5’-TACTTCCAGCACCGCACTCACATGCCCACATCC- 3’
gA6-3-2	5’-GGTGATGGGCAGCCACTTCGG-3’
gA7-3-1	5-’ACCGCACTCACGTGCCCACATCCACCATTCAAC- 3’
pdp45951-3-7	5’-GGGCAGTGCCGACTCAGCGAA-3’
pdp45951-3-8	5’-GGAGACTGACCGAACCTCTTCTC-3’
pdp45977-3-1	5’-ACTCAACAGTTTATCCCGCACTG-3’
pdp45977-3-3	5’-GTAAGCCAGCCACTCCAGACAC-3’
pdp45977-3-4	5’-CTCGACGGAGGCTTCTTCTTCCCG-3’
pdp45977-3-5	5’-GGACGTAGCACACCCTCTTC-3’

Chapter 3 Expression analysis of the PpCesA6 gene

3.1 Introduction

PpCesA6 and PpCesA7 are highly similar to each other, both on the cDNA and genomic DNA levels. There are only three amino acid differences between the proteins they encode. These two genes may arise from recent gene duplication. In theoretical population genetics, subfunctionalization plays an important role in stably maintaining the duplicated genes (Lynch and Force, 2000). It is believed that both duplicated genes can be stably maintained when they have some differences in their functions (Nowark *et al.*, 1997). The duplicated genes usually show divergence in expression (Zhang, 2003).

By taking advantage of homologous recombination of *Physcomitrella*, a reporter gene (such as GFP, YFP and GUS) can be fused with a gene of interest and targeted into the genome. The reporter gene can fuse with either C-terminus or N-terminus of the gene of interest. By using knocked-in gene targeting technique, the reporter gene can integrate into the genome and the expression of the reporter gene will be driven by the native promoter of the gene of interest. The strategy of targeted knock-in is shown in Figure 3-1.

The goal of this study is to investigate when and where PpCesA6 is expressed during the *Physcomitrella* life cycle by fusing a GFP expression cassette with this gene.

3.2 Materials and methods

GFP expression construct

Vector pGFPmutnptII was used for GFP expression construct (Figure 3-2). Forward primer GFP45951-F1-KpnI 5'-AATGGTACCAGTGGGCATCGACGAGTGG TG-3' and reverse primer GFP-45951-R1-HindIII 5'-AATAAGCTTAGCACCAGCACC ACAGTTTATCCCGCA-3' were used to amplify a 519 bp fragment that ended at the stop codon. The sequences of the added restriction sites were underlined. This fragment was inserted into the 5' multicloning site between the KpnI and HindIII sites of pGFPmutnptII. Forward primer GFP45951-F2-BamHI 5'-AATGGATCCGGGTGTGTG ATGGAGGAATCG-3' and reverse primer GFP45951-R2-NotI 5'-AATGCGGCCGCAC ACGTCTACCAAAGTCTC-3' were used to amplify a 431 bp fragment that is located just after the stop codon of PpCesA6. The sequences of the added restriction sites were underlined. This fragment was inserted into the 3'multicloning site between the BamHI and NotI sites of pGFPmutnptII. The resulting construct was named GFPCesA6 (Figure 3-3). The TOPO10 *E. coli* strain (Invitrogen; Carlsbad, CA, USA) was used in vector transformation.

PEG-mediated transformation

The resulting construct was amplified in *E. coli* and then linearized by double digestion with KpnI and NotI. The digested construct was extracted after gel electrophoresis. GFPCesA6 construct was also amplified by PCR with primers GFP-45951-F1-KnpI and GFP-45951-R2-NotI. Both restriction enzyme digested construct and PCR amplified construct were used for DNA transfer. 10-15µg DNA was used for

each transformation.

Protonemata were sub-cultured on BCDAT medium overlaid with cellophane. Protoplasts obtained from 4-6 days old sub-cultured protonemata were used for transformation. 3 plates of protonemata were digested in 10ml 1% Driselase (Sigma, D9515) solution prepared in 8% (w/v) mannitol for 1 hour at room temperature with gentle shaking. The protoplasts were filtered through a 100 µm nylon mesh, followed by centrifugation in a clinic centrifuge at 180×*g* for 3 minutes and resuspended in 10 ml 8% (w/v) mannitol. This procedure was repeated after the protoplast resuspension was filtered through a 40µm nylon mesh. The density of the protoplasts was counted with a hemocytometer. After the protoplasts were counted, they were resuspended in MMM solution (9.1% mannitol, 0.1% [w/v] MES [PH5.6], 15mM $CaCl_2$) at a concentration of 1.6×10^6/ml. 300µl of the protoplast suspension and 300µl of PEG solution (4g PEG6000 dissolved in a solution containing 9ml 8% [w/v] mannitol, 1ml 1M $Ca(NO_3)_2$ and 100µl 1M Tris-HCl[PH8.0]) were added to DNA solution(up to 30µl). After being heat shocked at 42^0C for 5 minutes, the mixed solutions were cooled at room temperature for 10 minutes. The mixture was then diluted with CaPW solution (containing 8% mannitol [w/v] and 50 mM $CaCl_2$) every 3 minutes, adding 300µl, 600µl, 1ml, 2ml, 3ml, 4ml, respectively. Afterwards, the suspension was centrifuged at 180×*g* for 3 minutes. The pellet was resuspended in 1 ml 8% [w/v] mannitol and then was mixed with 7ml protoplast regeneration medium top layer (PRMT) (BCDAT medium supplemented with 10mM $CaCl_2$, 8% [w/v] mannitol and 6% agar [w/v]). The mixture was poured into 4 10cm Petri dishes containing protoplast regeneration medium bottom layer (PRMB)

(BCDAT medium supplemented with 10mM $CaCl_2$, 6% [w/v] mannitol and 10% [w/v] agar) and sealed with 3M surgical tape. PRMT medium can be made in advance and melted PRMT can be kept at 45^0C in a water bath. The protoplasts were regenerated under continuous light for 6 days and then were transferred to BCDAT medium containing 50µg/ml G418 for two weeks. The surviving protoplasts were then transferred to BCDAT medium without antibiotics for two weeks. After that, a second period of selection of two weeks on BCDAT medium containing 50µg/ml G418 was followed (Grimsley *et al.*, 1977; Knight *et al.*, 2002; Nishiyama *et al.*, 2000; Schaefer *et al.*, 1991). The knockout DNA can form high molecular weight concatemers and can exist as extrachromosomal elements. Unstable clones exist at very high frequencies after transformation. They can be propagated for years as protonemata as long as the selective pressure is present. They lose their resistance if they are cultured on no-selective medium for a two weeks period (Ashton *et al.*, 2000; Schaefer, 2002). The second period of selection is necessary to ensure that the transformants are stable.

Molecular analysis of transformants

PCR-based analysis was used. Primers derived from the sequences located on the 5'genomic region outside of the fragment (A1) used in construct and the 5' side of the GFP expression cassette were used to check if the construct had integrated at the 5' locus.

Primers derived from the sequences located on the 3' side of the nptII selection cassette and the 3' genomic region outside of the fragment (A2) used in the construct were used to check if the construct had integrated at the 3' locus.

The forward primer KSCesA6-5-1 5'-TAACCTGGCGAGTCTGTGGTTTAT-3' is located 78bp upstream of where A1 starts. Forward primer KSCesA6-5-2 5'AATGC ACAGAGGATGGAGGTC-3' is located 513bp upstream of where A1 starts. Reverse primer GFP-R1 5'-TCACGAACTCCAGCAGGACCATGT-3' was based on the sequence at the 3' end of the GFP cassette. Forward primer sp-35ST-2 5'-GGTATC AGAGCCATGAATAGGTC-3'was based on the 35S terminator of CaMV. Reverse primer KSCesA6-3-1 5'-GACTATCTTGAAACCACACCGACA-3' was based on the sequence 244bp downstream of A2. Reverse primer KSCesA7-3-1 was based on the 3'UTR sequence of PpCesA7 outside the region that overlaps with PpCesA6. The sequence of KSCesA7-3-1 was 5'-GCACCTTCACTAATGCCCGACATA-3'.

Primesr GFP-A6-F1 and GFP-A7-F1 were derived from the first intron of PpCesA6 and PpCesA7, respectively. The sequence of GFP-A6-F1 is 5'-CACATCATGCAGCAAGACCCCCGA-3'. The sequence of GFP-A7-F1 is 5'-AAGTTCGCCATTCAGCTGCAGATATAC-3'. Thses two primers were used with GFP-R1 to verify that the GFP fusion construct was fused with PpCesA6.

DNA extraction

A quick DNA extraction method was used in screening the large amount of transformants. A small piece of protonemal tissue or a few gametophores were picked up from the culture medium with a pair of forceps. The tissue was smashed in a 1.5ml microcentrifuge tube containing 30µl 1×PCR buffer and was frozen in liquid nitrogen. The mixture was thawed at room temperature and was frozen in liquid nitrogen again.

The extract was then incubated at 65^0C for 10 minutes and was centrifuged at 14,000rpm for 5 minutes later. 2µl extract was used for PCR.

Induction of antheridia, archegonia and sporophytes

Protonemata were cultured on BCD medium at 25˚C under long day conditions (16 hours light and 8 hours dark) to promote differentiation. Gametophores with more than 10 leaves were transferred to short day conditions (8 hours light and 16 hours dark) at 16˚C to induce antheridia, archegonia and sporophytes (Hohe *et al.*, 2002).

Light Microscopy

The Leica DM IRBE fluorescence with a GFP filter was used to examine the GFP expression of GFP fusion lines. The Leica SP2 AOBS confocal laser scanning microscope was used to observe the cellulose synthase protein on the plasma membrane. The images were recorded in three channels (Core Facility, the University of Texas at Austin).

3.3 Results

3.3.1 GFP expression construct

When constructing the vector for homologous recombination, two different sizes of genomic DNA fragments from different regions of the gene were amplified with suitable restriction sites added to the primers and cloned into the multicloning sites of the backbone vector. The resulting construct was the selective marker expression cassette flanked by the moss sequences. The orientation of the selection marker gene complied with the orientation of the target gene.

In this study, a C-terminal GFP fusion construct was made. The vector pGFPmutnptII (kindly provided by RIKEN Bioresource Center, Japan) was used for the GFP fusion construct (Figure3-2). In this vector, the GFP coding region was followed by the nopaline polyadenylation signal (Tnos) as terminator and an nptII cassette was used as selection marker. The start codon of the GFP coding region was modified into TTG. The nptII cassette is driven by the CaMV 35S promoter and terminated by the CaMV polyadenylation signal. The nptII gene gives the transformants resistance to G418. The backbone of pGFPmutnptII was pBluescript II SK+, which had ampicillin resistance in *E. coli*.

A fragment (A1) of 519 bp of PpCesA6 cDNA (from position 2954 to position 3476), whose 3'end is the stop codon, was amplified by PCR with forward primer GFP45951-F1-KpnI and reverse primer GFP-45951-R1-HindIII. KpnI and HindIII restriction sites were added to the 5'and 3' primer, respectively. The stop codon was changed to GGT and a five-amino-acids-linker sequence (GCT GGT GCT AAG CTT), a string of Ala-Gly-Ala-Lys-Leu, was added to the 3' primer. This fragment was cloned into the 5' multicloning site between the KpnI and HindIII sites of pGFPmutnptII and was inserted before the start codon of the GFP sequence (which had already been modified from ATG into TTG). This fragment is part of the last exon of both PpCesA6 and PpCesA7 genomic DNA. Another fragment (A2) of 431 bp whose 5' end is just after the stop codon of PpCesA6 cDNA (from position 3455 to position 3886) was amplified by PCR with forward primer GFP45951-F2-BamHI and reverse primer GFP45951-R2-NotI. BamHI site was added to the 5'primer and Not I site was added to the 3'primer.

This fragment was inserted into the 3' multicloning site between the BamHI and NotI sites of pGFPmutnptII. This fragment also is part of the last exon of PpCesA6 genomic DNA. The first 140bp of this fragment is part of the last exon of PpCesA7 genomic DNA as well. The resulting construct was the GFP expression cassette and the nptII selection cassette flanked by two different PpCesA6 genomic DNA sequences. The orientation of the GFP and NptII cassettes complied with the orientation of the target gene. The resulting construct was named GFPCesA6. DNA sequencing was carried out to confirm the correct amplification of the homologous fragments and the correct cloning of these fragments, especially for the reading frame between A1 fragment and the GFP expression cassette. The overview of the GFPCesA6 construct is shown in Figure 3-3.

After the introduction of this construct into *Physcomitrella*, due to homologous recombination, the GFP cassette will fuse with the C-terminus of the protein. The expression of the GFP cassette will be driven by the native promoter of PpCesA6.

3.3.2 Polyethylene glycol (PEG)-mediated protoplast transformation

The polyethylene glycol (PEG)-mediated DNA transfer method has been used routinely for gene targeting in *Physcomitrella* (Schaefer *et al.*, 1991; Hohe *et al.*, 2004). The developmental stage of tissue used for protoplast isolation, the osmolarity of the solutions for protoplast isolation and culture, and the PEG solution are critical factors for a successful transformation. Protonemata growing on the culture medium with ammonium have more chloronemata. Figure 3-4 shows the plant material used for protoplast isolation and the protoplasts and protonemata after transformation. Sub-

cultured 4-6 days old protonemata are the best tissue for protoplast isolation (Fig3-4A). Protoplasts isolated from older protonemata have lower regeneration ability than younger protonemata after being transferred with foreign DNA. PEG solution hydrolyzes at room temperature. This will change the PH of the solution and subsequently will affect protoplast regeneration. This solution should be made fresh on the day of transformation.

The construct can form concatemers and exist as the high molecular weight extrachromosomal element. Episomal antibiotic resistant clones have been obtained in every transformation. The extrachromosomal elements replicate in moss cells but they do not participate in mitosis (Ashton *et al.*, 2000; Schaefer, 2002). Two rounds of selections are needed in selecting stable transformants. The protoplasts are cultured on the non-selective medium for one week to facilitate protoplast regeneration. Then, the regenerated protoplasts are transferred to the culture medium with antibiotics and are cultured for 2 weeks. The surviving clones are transferred back to the non-selective medium for one to two weeks and then undergo the second selection.

3.3.3 Molecular analysis of GFP fusion lines

PCR-based analysis is used to screen transformants. Two sets of primers are used in PCR. One set anneals at the 5' side of genomic DNA outward of the homology that was used to make the construct and the 5'side of the selection marker. The other set anneals at the 3' side of the selection marker and the 3' side of the genomic DNA outward of the homology that was used to make the construct (see Figure 3-1). The purpose of the first set PCR is to check if the construct has integrated at the 5'end of the

targeted locus. The second set PCR is to check if the construct has integrated at the 3' end of the targeted locus.

PpCesA6 and PpCesA7 are very similar to each other. The homologous fragment A1 in GFPCesA6 construct exists in both PpCesA6 and PpCesA7. About 140bp of the 5'end of fragment A2 exists in both PpCesA6 and PpCesA7. When this construct inserts into the wild type genome, it may fuse with either one of these two genes or both at the same time. To determine which gene that the GFP expression cassette has fused with, several sets of PCR were carried out. Forward primers KSCesA6-5-1 and KSCesA6-5-2 were used with reverse primer GFP-R1 to check if the construct had integrated at the 5' end of the targeted locus. Forward primer sp-35ST-2 and reverse primer KSCesA6-3-1 or KSCesA7-3-1 were used to check the integration of construct at the 3' end of the targeted locus.

Transformation line g3-4-3-1 was the first identified GFP fusion line. When using KSCesA6-5-1 and GFP-R1 as primers, an expected 1269 bp product was obtained. When using KSCesA6-5-2 and GFP-R1 as primers, an expected 1708 product was obtained. When using primers sp-35ST-2 and KSCesA6-3-1, an expected 1256bp product was obtained (Figure 3-5A, B). These PCR products were purified after gel electrophoresis and were submitted for DNA sequencing. DNA sequencing results confirmed that the GFP expression cassette had fused with PpCesA6.

In order to check if the GFP construct had also integrated with PpCesA7, primers sp-35ST-2 and KSCesA7-3-1 were used for PCR analysis. A 700bp band was obtained (Figure 3-6). This result suggested that the nptII selection cassette had integrated with

PpCesA7. However, the length of the product was shorter than expected. DNA sequencing result showed that the integration site was at the end of the region (A2 on the construct) that overlaps with PpCesA6. The PCR product was 140bp shorter than expected. During homologous recombination, targeted insertion (in which homologous recombination only occurs in one arm of the knockout vector, accompanied by non-homologous end-joining (NHEJ) by the other arm of the vector) occurs at a certain frequency (Kamisugi *et al.*, 2006). This result indicated that targeted insertion occurred at the 3' end (Arm2) of the construct with PpCesA7. In order to rule out that the GFP expression cassette had fused with PpCesA7, two measurements were taken. First, PCR reactions with primers based on the first intron of both genes as forward primer and GFP-R1 as reverse primer were carried out. Primer GFP-A6-F1 and primer GFP-A7-F1 were designed based on PpCesA6 and PpCesA7, respectively. Second, the PpCesA7 knockout line (see chapter 4) was transformed with GFPCesA6 construct.

PCR reactions were carried out by using different DNA polymerases and the experiments were repeated. An about 6.7 kb size band was obtained in all reactions (Figure 3-7). The bands were purified from the gel after electrophoresis and submitted for DNA sequencing. The DNA sequencing result showed that the PCR product from reactions with GFP-A6-F1 and GFP-R1 as primers matched the sequence when the GFP expression cassette fused with PpCesA6. However, the PCR product from the reactions with GFP-A7-F1 and GFP-R1 as primers did not give any DNA sequencing result. This suggested that the GFP-A7-F1 and GFP-R1 PCR product was caused by mispriming. In the PpCesA7 knockout line transformed with the GFP construct, the transformants

observed with green fluorescence had the same expression pattern as the transformants derived from wild type transformed with GFP construct. PCR results also proved that the GFP expression cassette had fused with PpCesA6 (data not shown). All together, these results suggested that PpCesA6 was the only gene that was fused with the GFP expression cassette.

Several other GFP fusion lines were identified by the same method. All these lines were used for further study.

3.3.4 Microscopy

GFP fusion lines were observed under fluorescence microscope with GFP fluorescence filter. Green fluorescence was observed in both protonemata and young gametophore (Figure 3-8 and Figure 3-10, 11, 12). In young gametophore, green fluorescence was observed in axillary hairs and basal rhizoid cells. The GFP fusion lines were also cultured at 16^{o}C under short light cycle to induce gametangia and sporophyte. No green fluorescence was observed in the gametangia and the sporophyte.

In protonemata, green fluorescence was observed in both chloronema and caulonema cells, but not in every cell (Figure 3-8, 3-9). In order to determine the specific time when PpCesA6 was expressed, protoplasts of the g3-4-1 fusion line were isolated and cultured on suitable medium. The regenerated protoplasts were examined with the microscope every week. Green fluorescence was detected as early as two weeks after protoplast regeneration. Overall, not every protonema cell showed green fluorescence. The apical cell of protonema had the strongest green fluorescence signal.

In the rhizoid, green fluorescence was detected as early as these cells started to grow, even before gametophore development (figure 3-13). In developing gametophores, green fluorescence was observed in initiate rhizoid cells as well as mature rhizoid cells (Figure 3-10, 3-12).

Protonema cell from g3-4-31 was observed by laser scanning microscope. The green fluorescence was detected in the focal plane of the plasma membrane, showing that CesA protein is localized on the plasma membrane. The green particles were randomly dispersed (Figure 3-14).

3.4 Discussion

Studies of cellulose synthase genes in other systems have shown that the expressions of CesAs are very tissue and cell-specific. For example, ZmCesA1 and 6-8 are mainly responsible for primary cell-wall biosynthesis in the stalk. ZmCesA6 has the highest expression in leaves and is the most highly expressed CesAs in this tissue. Among the cell division zone, elongation zone and transition zone of a developing leaf, the expression of ZmCesA6 is very low in the division zone and very high in the transition zone (Appenzeller *et al.*, 2004).

By fusing a GFP expression cassette with PpCesA6 at its C-terminus, the GFP expression was driven by the native promoter of PpCesA6. Green fluorescence was detected in both protonema and gametophore. Both chloronema and caulonema cells had GFP expression, but not all the protonemata cells had the GFP expression. This result suggests that PpCesA6 is only expressed at certain developmental stage of protonemata.

In gametophores, the green fluorescence was detected in the axillary hairs and basal rhizoids. These results suggest that PpCesA6 is differentially expressed in a tissue specific manner. Analysis of the relative abundance of CesA ESTs from *Physcomitrella* developing sporophytes library, chloronemata and young gametophores library, auxin-treated chloronemata, caulonemata, rhizoids library, cytokinin-treated chloronemata, caulonemata and buds library showed that PpCesA6 and PpCesA7 were expressed only in the auxin-treated library that were derived from chloronemata, caulonemata, rhizoids cultures and cytokinin-treated library that derived from chloronemata, caulonemata and bud cultures (Roberts and Bushoven, 2007). Since both libraries had protonemata and rhizoid cells, my study is consistent with these results.

Rhizoids are important structures in the non-vascular plants that do not have roots. They have similar functions as roots. In *Physcomitrella*, rhizoids are brown-pigmented multicellular filaments. They develop from the epidermal cells of the stem. There are two types of rhizoids in *Physcomitrella*: basal and mid-stem rhizoids. Basal rhizoids appear first during gametophore development. They are at the base of the gametophore, below the first leaf. Mid-stem rhizoids arise from the stem of adult gametophore, between the leaves (Sakakibara *et al.*, 2003).

Axillary hair cells are used in taxonomic observation in moss (Griffin and Buck, 1989; Hedenas, 1990). They are located at the adaxial side of leaves. Their numbers vary from one to several per leaf axil. They secrete mucilage to prevent dehydrating in juvenile leaves (Ligrone, 1986; Shaw and Goffinet, 2000). There was no report about

how many cells are in each hair in *Physcomitrella*. This study showed that there are two cells in each hair in *Physcomitrella* (Figure 3-15).

Trichomes and root hair cells in *Arabidopsis* have the same lineage. Molecular genetic studies have shown that many genes regulate the patterning of both trichomes and root hair cells (Pesch and Hulskamp, 2004; Ojangu *et al.*, 2007). Rhizoid cells and axillary hairs all arise from the stem. It is not surprising that PpCesA6 is expressed in these two specific cells at the same time.

Vascular plants have both primary and secondary cell walls. In *Arabidopsis*, five CesAs are involved in primary cell wall synthesis and another three CesAs are involved in secondary wall synthesis (Desprez *et al.*, 2002, 2007; Fagard, et al, 2001; Taylor *et al.*, 2000, 2003). Three CesAs in rice, three CesAs (ZemCesA10-12) in maize and three CesAs in aspen (PtrCesA1-3) are involved in secondary wall synthesis (Appenzeller *et al.*, 2004; Joshi *et al.*, 2004; Tanaka *et al.*, 2003). Many of these genes show very cell-specific expression. For example, Zm CesA8 is expressed differentially in epidermis at the root tip. AtCesA6 is involved in primary cell wall cellulose synthesis in dark-grown hypocotyls and in the roots (Fagard *et al.*, 2000). All the cells in *Physcomitrella* only have primary cell wall. There are almost as many cellulose synthase genes in *Physcomitrella* as in *Arabidopsis*. Thus, it is not surprising that the expression of CesAs in *Physcomitrella* is more tissue-specific.

Cellulose synthase proteins are transmembrane proteins (Kimura *et al.*, 1999). Confocal microscope study showed that PpCesA6 was localized on the plasma membrane. A confocal microscope study also showed that the protein particles were

randomly dispersed on the cell membrane. Cortical microtubules and cellulose fibrils are frequently observed in parallel to each other. It is believed that the cortical microtubules can guide the cellulose synthase complex (Baskin, 2001). Visualization of cellulose synthase complex in *Arabidopsis* expanding hypocotyl demonstrated that: (1) the cortical microtubules are roughly transverse to the cell axis; (2) the cellulose synthase complexes and microtubules are indeed connected; and (3) the cellulose synthase complexes follow the same paths as microtubules (Paredez *et al*., 2006). Immunofluorescence microscopy study of microtubules in *Physcomitrella* chloronema and caulonema cells showed that cortical microtubules form a three-dimensional meshwork (Doonan and Cove, 1985). Since the cortical microtubules of *Physcomitrella* are dispersed randomly, it is not surprising that cellulose synthase complexes are also dispersed randomly.

PpCesA6 and PpCesA7 genes encode almost the same protein. It is very possible that PpCsA7 is expressed in the same cells as PpCesA6. If the genes are expressed in unique cell types, it is difficult to accurately study the expression patterns associated with development by using northern, RT-PCR analysis etc.

3.5 Acknowledgements

I would like to thank Dr. Rumiko Kofuji at the National Institute for Basic Biology, Myoudaiji-cho, Okazaki-shi, Japan for giving me permission to use pGFPmutnptII vector and RIKEN Bioresource Center, Tsukuba-shi, Ibaraki, Japan for providing me with this vector. I would also like to thank Dr. Angela Bardo and the core facility of the University of Texas at Austin for the microscopy work.

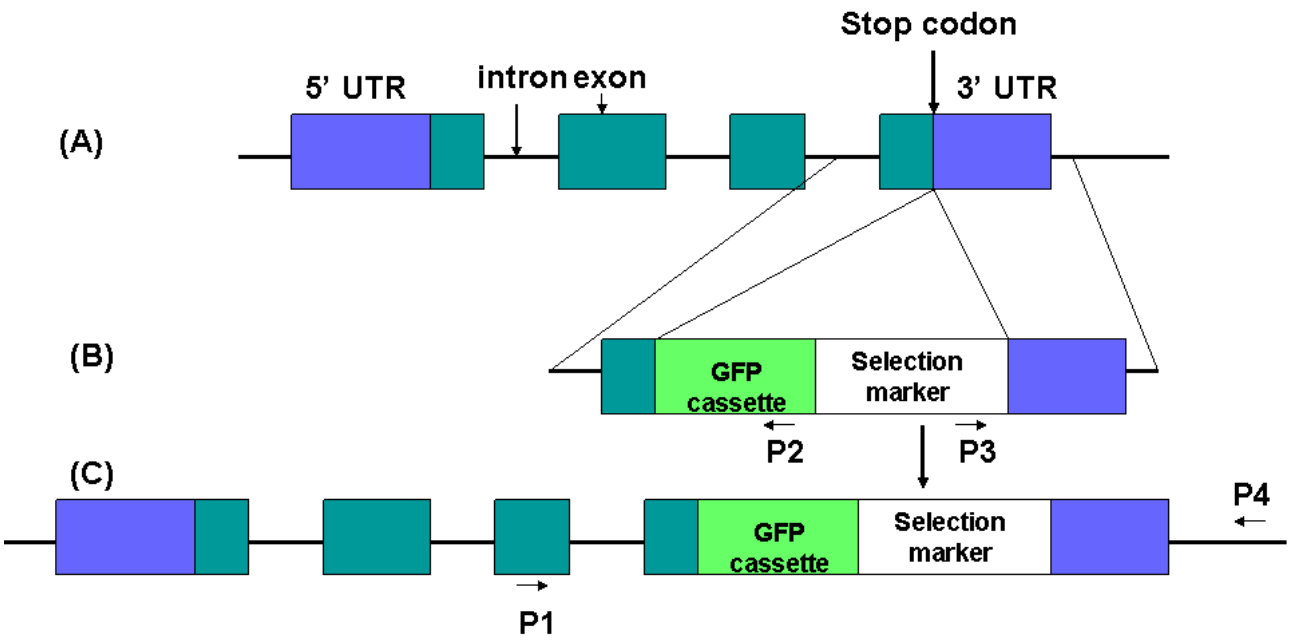
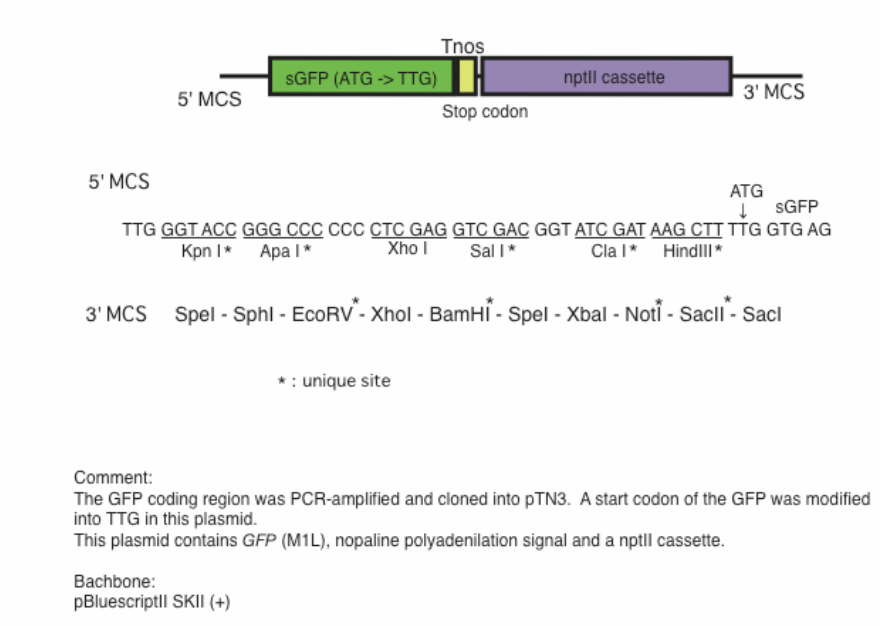


Figure 3-1 Strategy of targeted knock-in

(A) Structure of the gene of interest. Exons are in dark green squares. UTRs are in blue squares. Lines between squares indicate introns.
(B) Structure of knock-in construct. On the left side of the GFP cassette and the selection marker is the first homologous fragment that ends at the stop codon. On the right side of the GFP cassette and the selection marker is the second homologous fragment that is located just after the stop codon.
(C) The outcome after homologous recombination.

Two sets of primers are used for PCR based analysis. Primers P1 and P2, Primers P3 and P4 are used together, respectively.



Comment:
The GFP coding region was PCR-amplified and cloned into pTN3. A start codon of the GFP was modified into TTG in this plasmid.
This plasmid contains *GFP* (M1L), nopaline polyadenilation signal and a nptII cassette.

Bachbone:
pBluescriptII SKII (+)

Figure 3-2 Structure of pGFPmutnptII vector

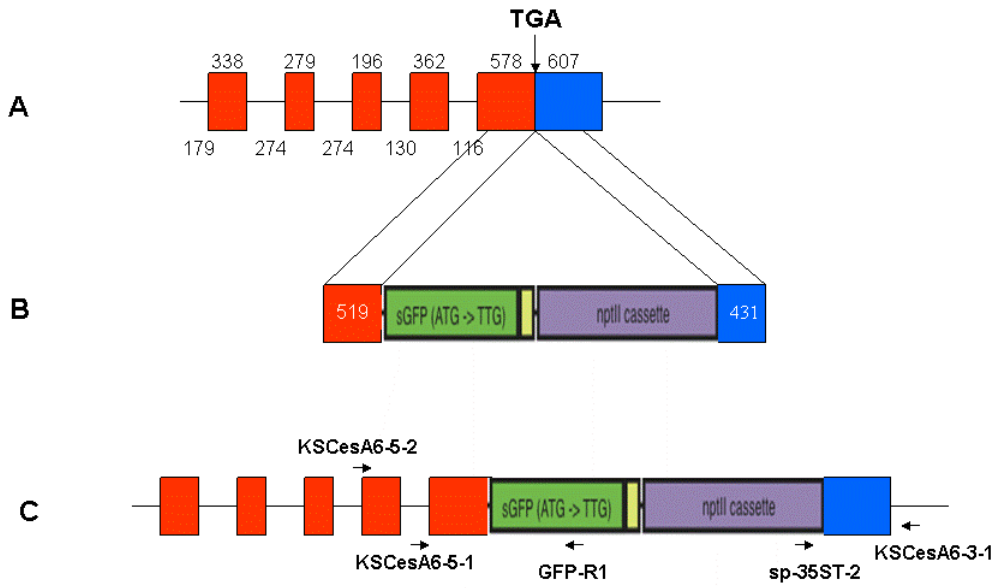
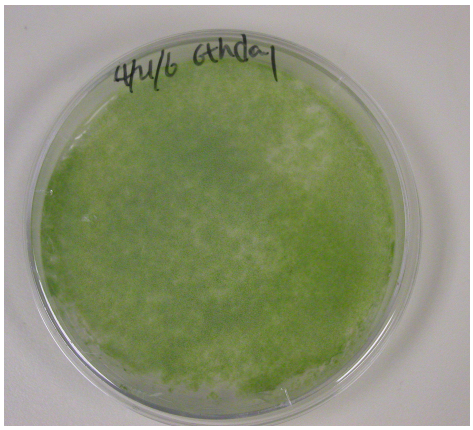


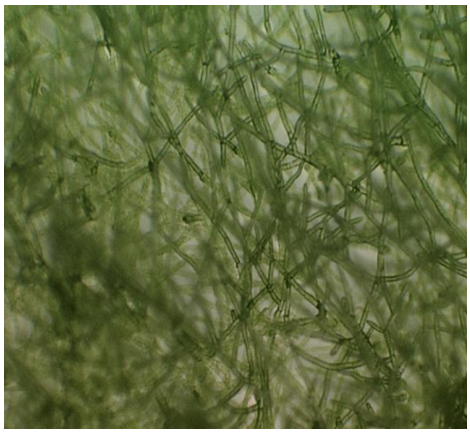
Figure 3-3 Strategy of PpCesA6 GFP fusion construct

(A) Gene structure of PpCesA6. The last four exons and 3'-UTR are shown.
(B) Structure of GFPCesA6 construct. The red square represents the sequence before the stop codon TGA. The blue square and the black line represent the sequence after the stop codon TGA .
(C) The outcome after targeted knock-in.

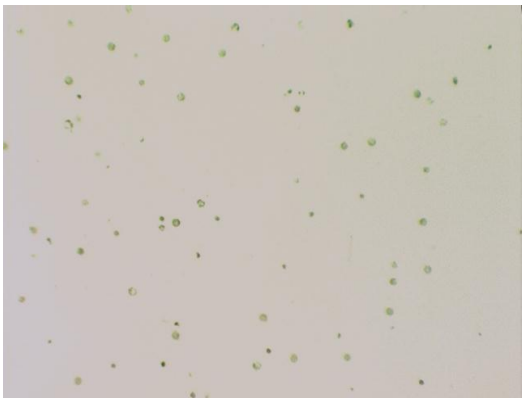
Primers KSCesA6-5-1 and KSCesA6-5-2 were used with GFP-R1 in PCR analysis to check if the construct had integrated at the 5' of the targeted locus.
Primers sp-35ST-2 and KSCesA6-3-1 were used in PCR analysis to check if the construct had integrated at the 3' of the targeted locus.



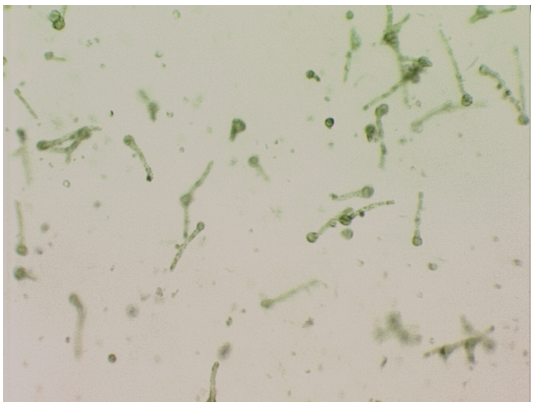
A



B



C



D

Figure 3-4 Protonemata used for protoplast isolation and the protoplasts and protonemata after transformation

A, B 6-day-old protonemata
C protoplasts-24 hours after transformation
D protonemata-6 days after transformation

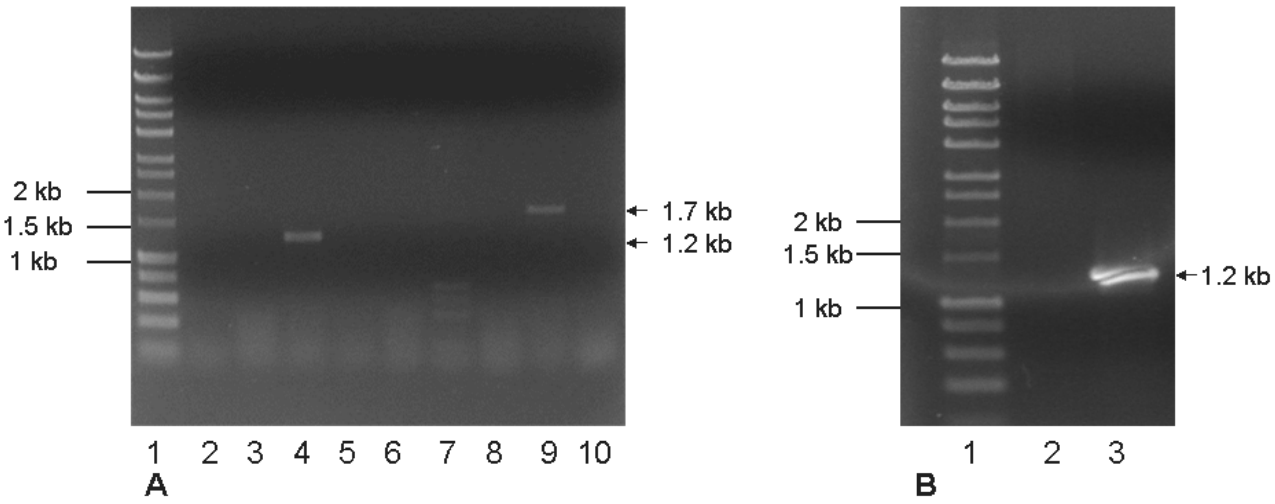


Figure 3-5 PCR analysis of GFP fusion lines

A Lane1, DNA marker.
Lane 2-6, KSCesA6-5-1 and GFP-R1 as primers.
Lane2, wild type. Lane 4, g3-4-3-1.

Lane 7-10, KSCesA6-5-2 and GFP-R1 as primers.
Lane 7, wild type.
Lane9, g3-4-3-1.
Arrows indicate the 1.2kb and 1.7kb products.

B sp-35ST-2 and KSCesA6-3-1 as primers.
Lane 1, DNA marker. Lane 2, wild type. Lane3, g3-4-3-1.
Arrow indicates the 1.2 kb product.

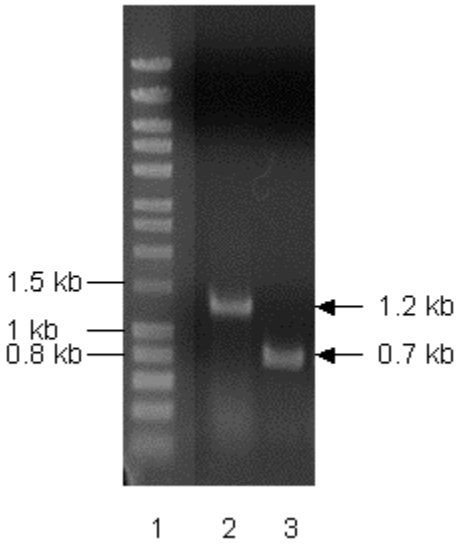


Figure 3-6 700bp PCR product of sp-35ST-2 and KSCesA7-3-1 from g3-4-3-1
Lane 1, DNA marker.
Lane 2, sp-35ST-2 and KSCesA6-3-1 as primers.
Arrow indicates the 1.2kb PCR product.
Lane 3, sp-35ST-2 and KSCesA7-3-1 as primers.
Arrow indicates the 0.7kb PCR product.

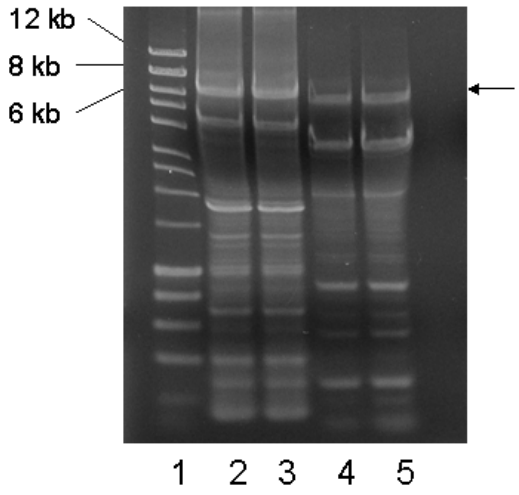


Figure 3-7 6.7kb PCR product from g3-4-3-1
Lane1, DNA marker.
Lane2 and 3 GFP-A6-F1+GFP-R1 as primers.
Lane4 and 5 GFP-A7-F1+GFP-R1 as primers.
Arrow indicates the 6.7 kb PCR product.

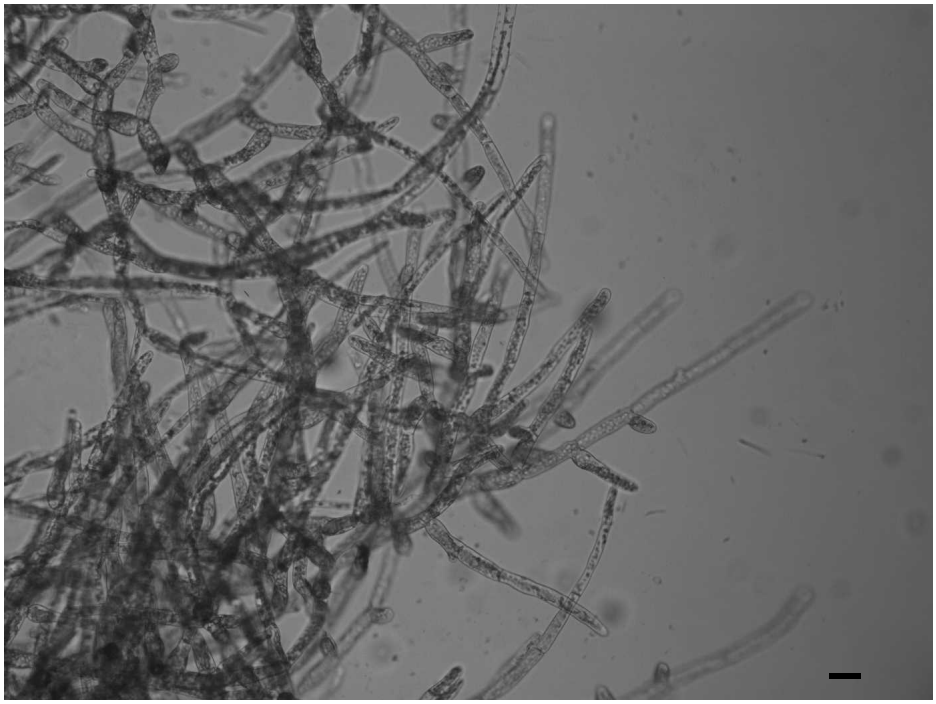


Figure 3-8 A
Bright field image of protomemata of g3-4-3-1
Scale bar=50µm

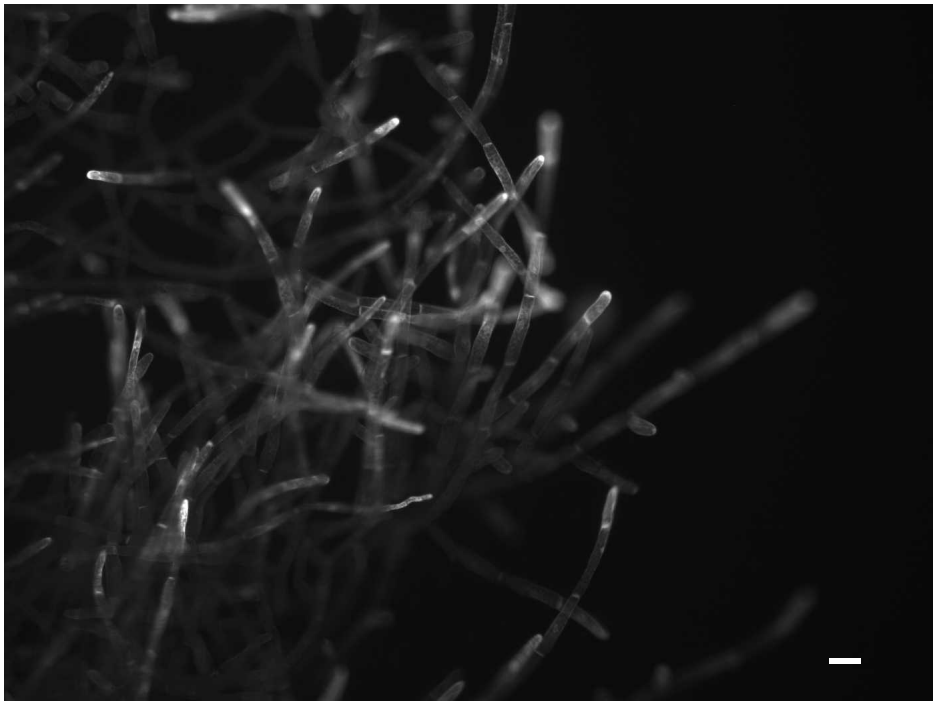


Figure 3-8 B
Green fluorescence image of protonemata of g3-4-3-1
Note that apical cells have the strongest fluorescence Scale bar=50µm

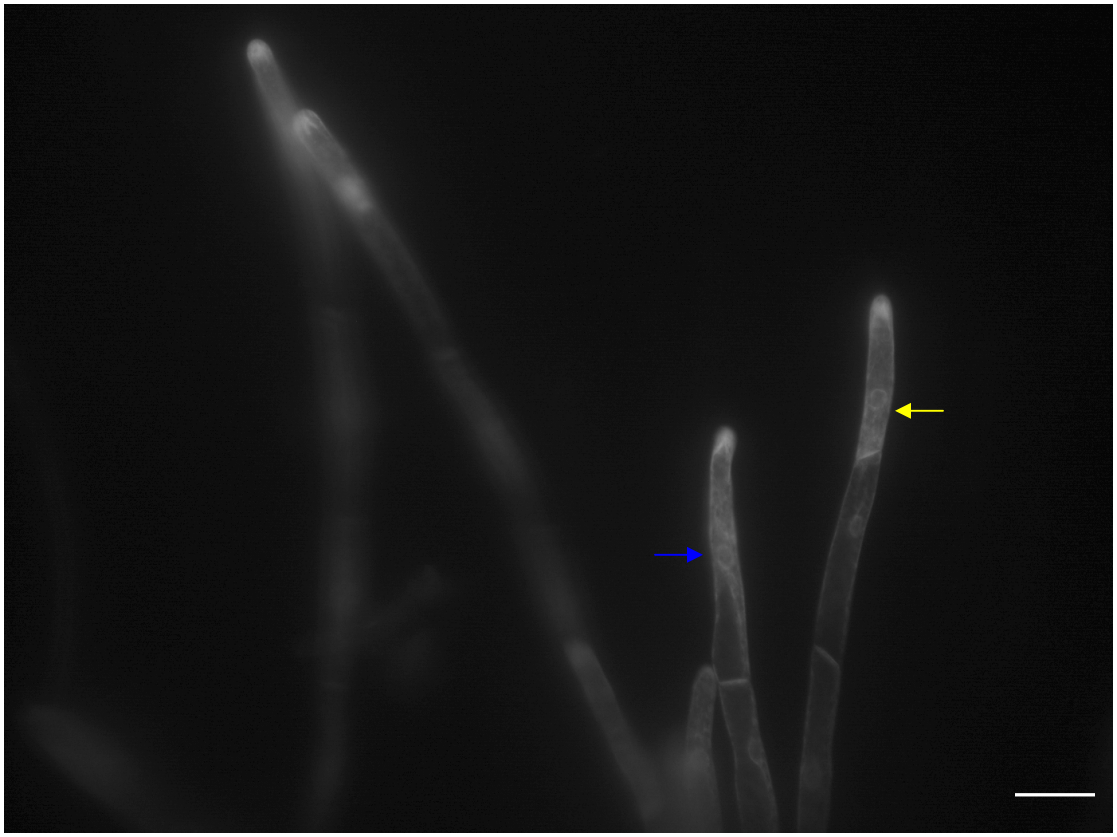


Figure 3-9 Image of green fluorescence in chloronema and caulonema cells

Blue arrow indicates a chloronema cell

Yellow arrow indicates a caulonema cell

Note the transverse cell wall in chloronema cell and oblique cell wall in caulonema cell.

Also Note that the apical cells have the strongest green fluorescence signal.

Scale bar = 50μm

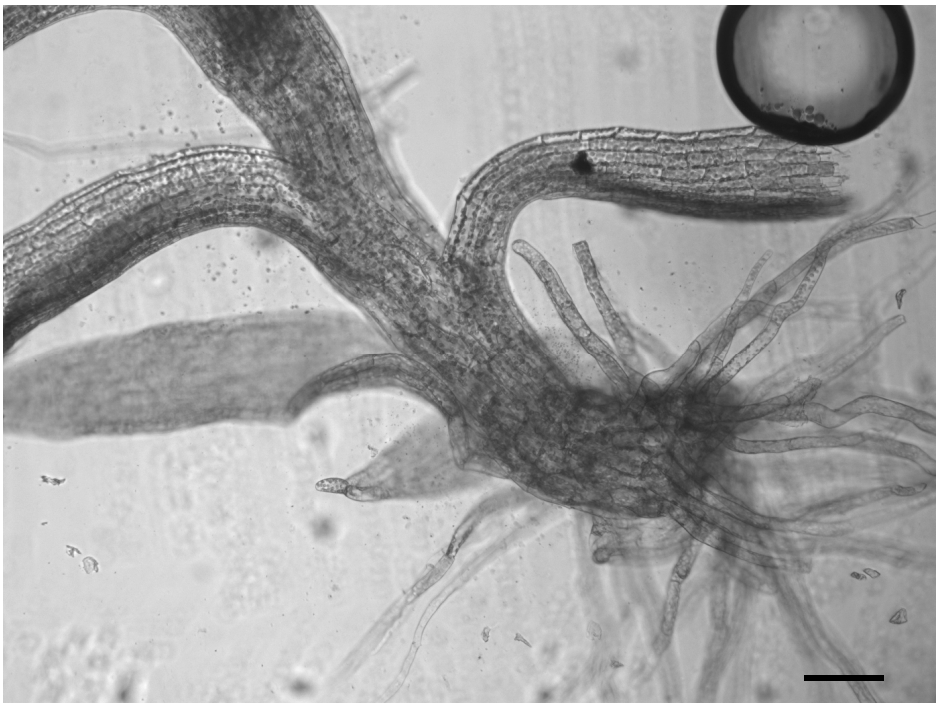


Figure 3-10 A
Bright field image of a gametophore of g3-4-3-1 scale bar=100µm

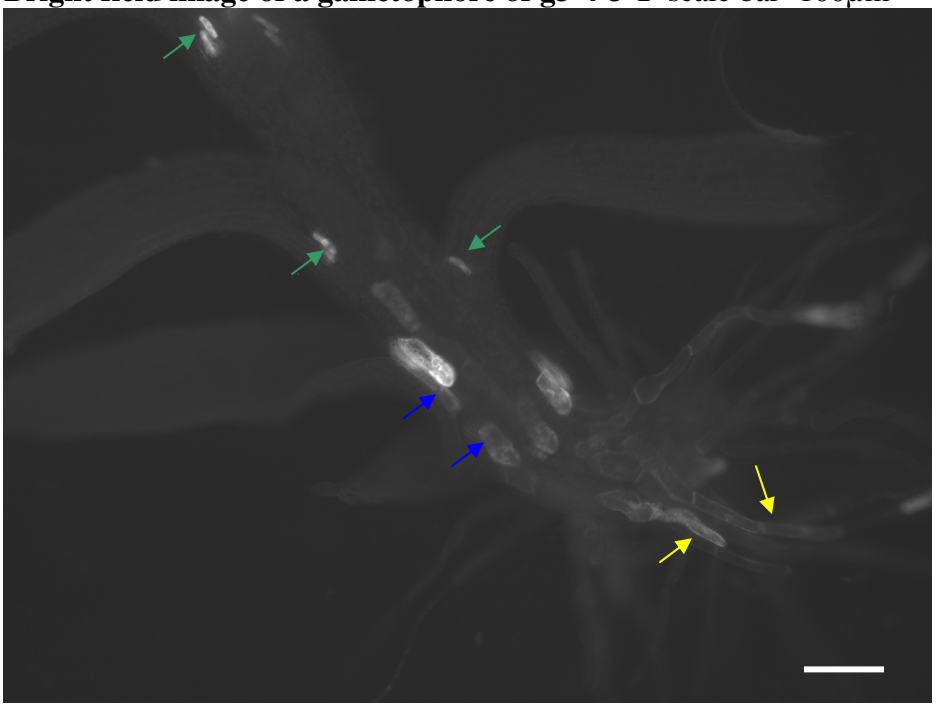


Figure 3-10 B Green fluorescence of a gametophore of g3-4-3-1
Note the fluorescence in axillary hair cells (green arrows), initiate rhizoid cells (blue arrow) and rhizoid cells (yellow arrow). scale bar=100µm

Figure 3-11 A
Bright field image of part a gametophore of g3-4-3-1 (lower magnification) scale bar=100µm

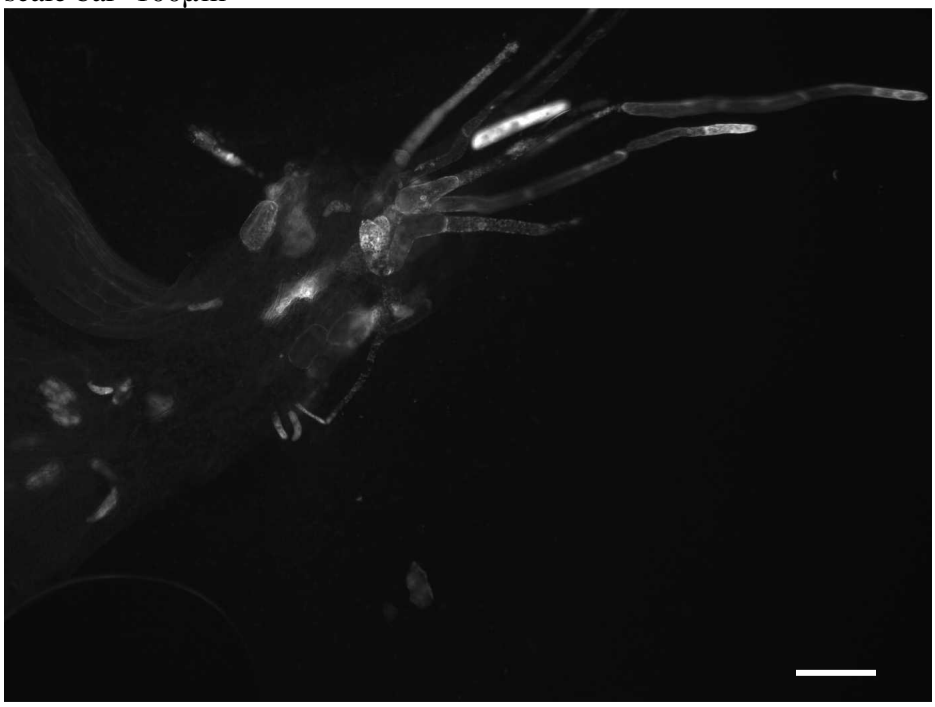


Figure 3-11 B
Green fluorescence image of A (lower magnification) scale bar=100µm

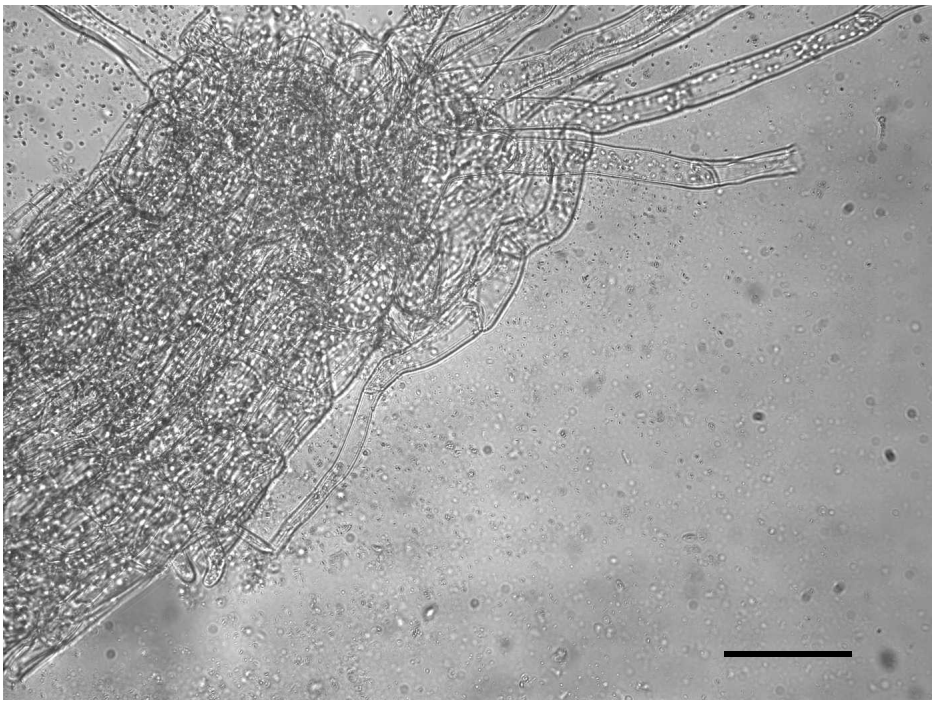


Figure 3-12 A Higher magnification of the bright image of Figure 3-11 A
Scale bar=100µm

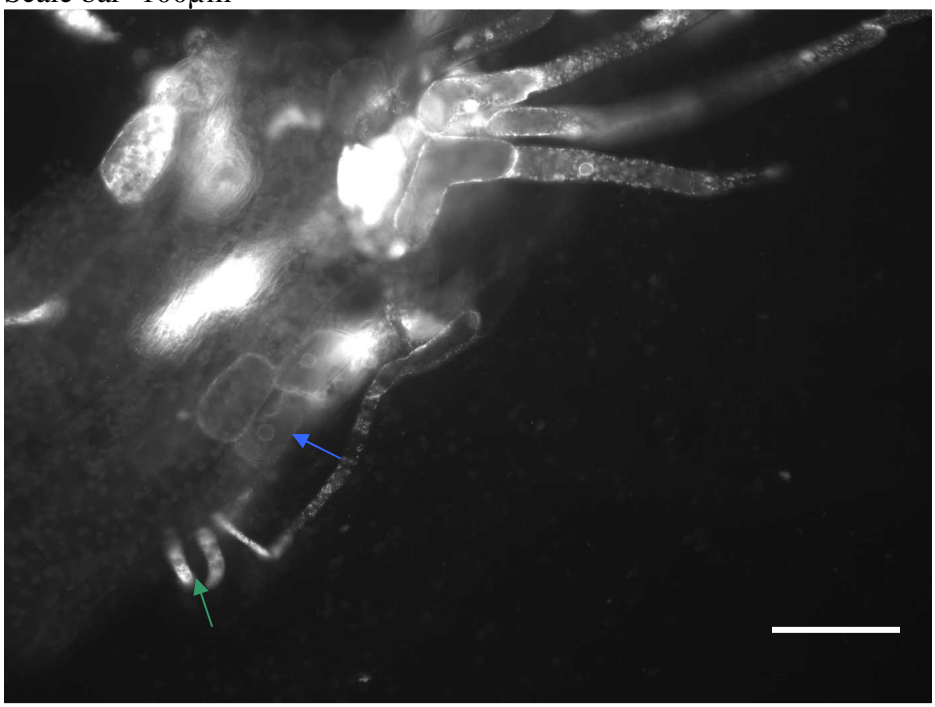


Figure3-12 B Higher magnification of the image of Figure 3-11 B
Note the two hair cells (green arrow) and initiate rhizoid cells (blue arrow)
Scale bar=100µm

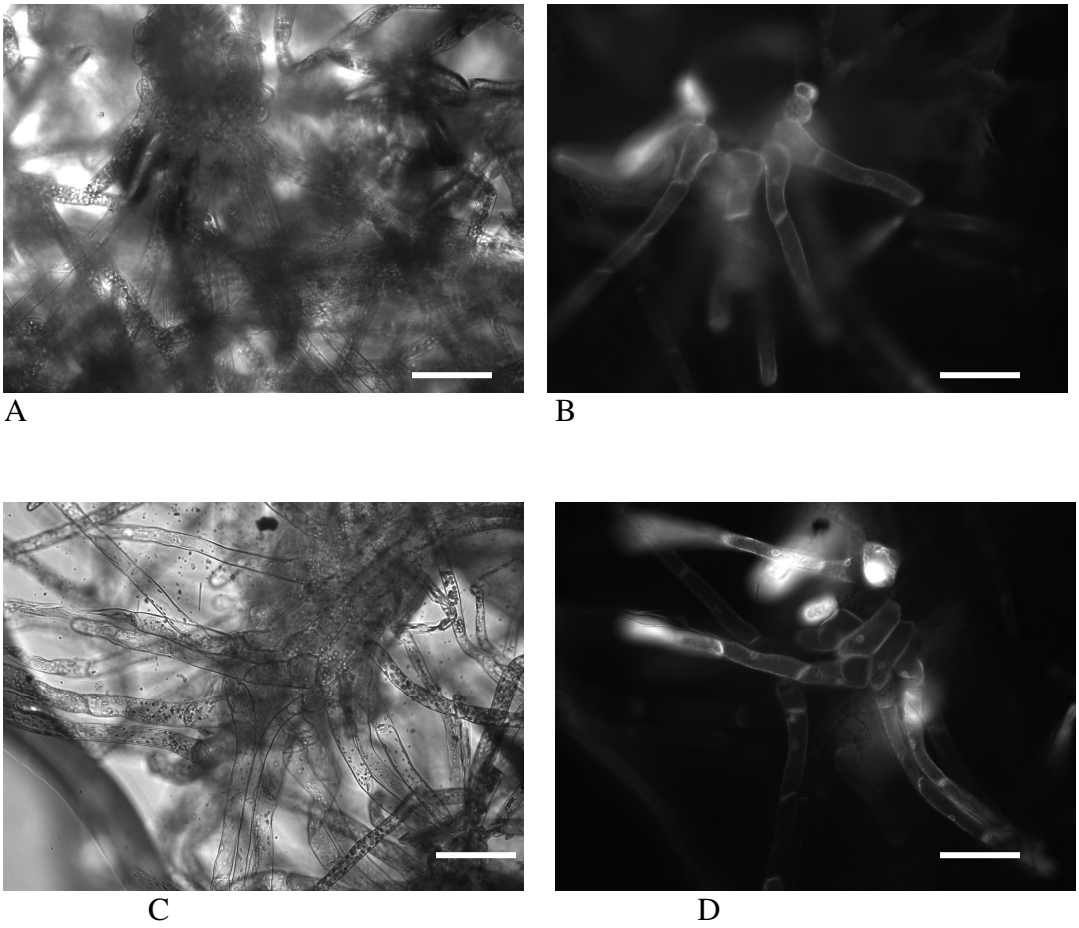


Figure 3-13 GFP expression in rhizoid cells

A, B images of line g3-4-5-2
C, D images of line g3-4-5-6
A, C bright field images
B, D GFP fluorescence images
Note the rhizoid initiate cells and mature rhizoid cells.
Scale bar=100μm

Figure 3-14 A Bright field image of a chloronema cell

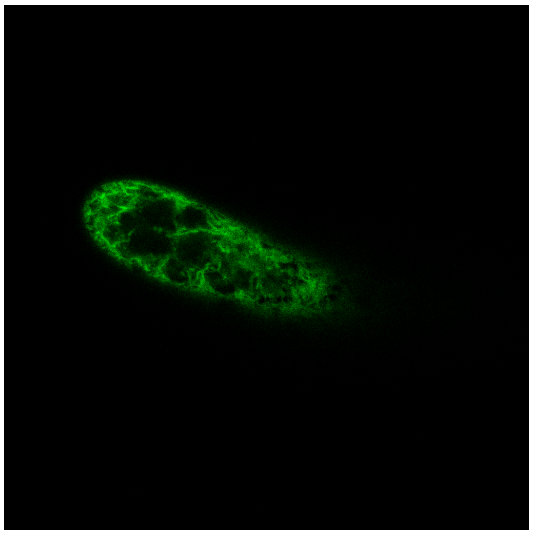


Figure 3-14 B Green fluorescence image of A

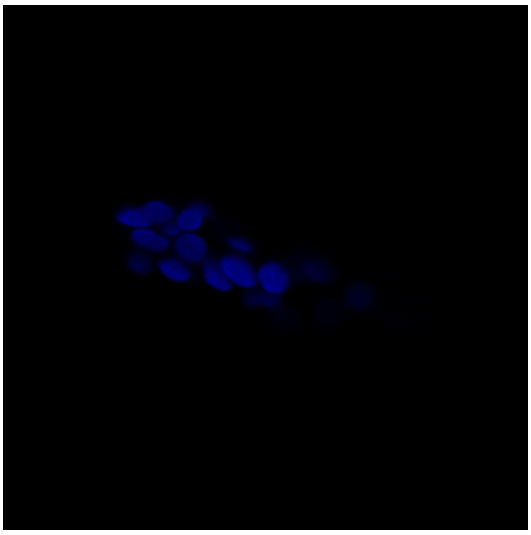


Figure 3-14 C Image of chloroplasts

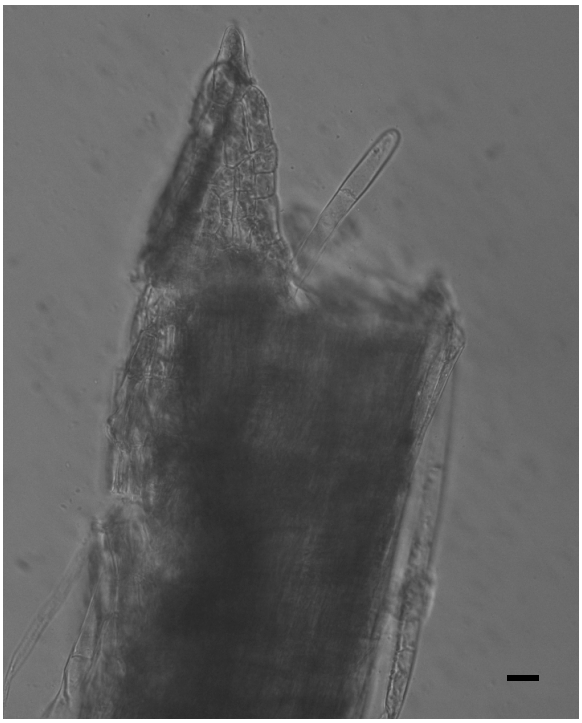


Figure 3-15 Bright field image of a axillary hair
Most of the stem and leaves have been removed to expose the hair.
Note that there are two cells in the hair.
Scale bar=20µm

Chapter 4 Targeted knockout of PpCesA6 and PpCesA7

4.1 Introduction

Physcomitrella patens has a very high frequency of homologous recombination (HR) in its nuclear genome (Schaefer and Zryd, 1997), whereas in flowering plants, HR occurs sporadically and at low frequency (Hanin *et al.*, 2001; Hanin and aszkowski, 2003). Because of this unique feature, genes of interest can be precisely targeted by exploiting homologous recombination in *Physcomitrella*. Targeted knockout is achieved by replacing the genomic DNA sequence with a selection marker cassette (figure 4-1). Gene targeting can be achieved by transforming the DNA construct that contains homologous sequences with genomic loci into the organism. The DNA construct used for transformation is the selection marker (such as the hygromycin and G 418 resistance markers) flanked by 2 different homologous regions (arms). A null mutant can be generated when the coding region of the targeted gene is replaced by the selection marker.

Both cDNA and genomic DNA sequences can be used for making targeting constructs. The target efficiencies of homologous regions based on cDNA sequences are less than those in which the genomic DNA sequences are used for making targeting constructs. The length of the homologous regions in DNA construct varies. Homologous regions as short as 300bp were used successfully in gene targeting. Research shows that the frequency of homologous recombination greatly depends on the length of homology.

Frequencies are generally low with shorter constructs. An overall homology of only 1 kb is sufficient to achieve a 50% yield of targeted transformants (Schaefer and Zyrd, 1997; Kamisugi *et al.*, 2005).

Gene targeting is usually performed by the polyethylene glycol (PEG)-mediated transformation of *Physcomitrella* protoplasts with the DNA construct. Both linear and circular construct DNA can be used for transformation (Schaefer *et al.*, 1991; Hohe *et al.*, 2004).

The goal of this study was to examine the role of PpCesA6 and PpCesA7 genes in cellulose biosynthesis. PpCesA6 and PpCesA7 gene disruptants were generated by homologous recombination. The null mutants of both genes were generated by replacing the coding region with a selection marker. PpCesA6 and PpCesA7 double mutants were also generated by using the same strategy.

4.2 Materials and methods

Gene knockout constructs of PpCesA6 and PpCesA7

For PpCesA6 knockout construct**,** the pTN3 vector containing an nptII cassette was used for producing the construct (Figure 4-2). The forward primer KOCesA6-1-F1-ApaI with an ApaI site and the reverse primer KOCesA6-1-R1-HindIII with a HindIII site were used to amplify the 1058bp fragment A1 containing the 5' UTR, the first exon and part of the first intron. The sequence of the forward primer KOCesA6-1-F1-ApaI is 5'- AATGGGCCCTTCTAATATTAGTGATGGCCT-3. The sequence of the reverse primer KOCesA6-1-R1-HindIII is 5'-AATAAGCTTAATTAAGAATGCTGAACATAG

-3. The sequences of the added restriction sites were underlined. This fragment was cloned into the 5' MCS of pTN3. The forward primer KOCesA6-F2-BamHI with a BamHI site and the reverse primer KOCesA6-R2-SacI with a SacI site were used to amplify the 1013bp fragment A2 containing the 3' UTR of PpCesA6 (starts from 113bp after the stop codon). The sequence of KOCesA6-F2-BamHI is 5'-AATGGATCCTTTC CGTGGTGCACGGAGGAA-3. The sequence of KOCesA6-R2-SacI is 5'-AATGAGC TCCACTGTCGACCAAAGGTGTGT-3. The sequences of the added restriction sites were underlined. This fragment was cloned into the 3'MCS of pTN3. DNA sequencing was carried out to confirm that there were no errors during amplifying and cloning the sequences.

For PpCesA7 knockout construct**,** the vector p35S-Zeo was used for producing the construct (Figure 4-5). The forward primer KOCesA7-A1-F1-KpnI with a KpnI site and the reverse primer KOCesA7-A1-R1-XhoI with an XhoI site were used to amplify the 980bp fragment A1 (containing the 5'UTR and part of the first exon, 60bp after the start codon) of PpCesA7. The sequence of KOCesA7-A1-F1-KpnI is 5'-TGCGGTACC AATAAGTTCAAGGATATCAGT-3. The sequence of KOCesA7-A1-R1-XhoI is 5'-TAACTCGAGAATAACAACCAGCTCGTTCC-3'. The sequences of the added restriction sites were underlined. This fragment was cloned between the KpnI and XhoI sites of p35S-Zeo. The forward primer KOCesA7-A2-F1-NotI with a NotI site and the reverse primer KOCesA7-A2-R1-SacI with a SacI site were used to amplify the 1032bp fragment A2 (containing the 3' sequence of PpCesA7 that starts 190bp after the stop codon). The sequence of KOCesA7-A2-F1-NotI is 5'-AATGCGGCCGCATTTGTTAA

CGGTGGTGT-3. The sequence of KOCesA7-A2-R1-SacI is 5'-GAGGAGCTCCTATC AAGCCTTGCAAGAT-3. The sequences of the added restriction sites were underlined. This fragment was cloned between the NotI and SacI sites of p35S-Zeo. DNA sequencing was also carried out to confirm that there were no errors in the amplified sequences and cloned sequences.

***Physcomitrella* protoplast transformation**

The knockout construct was digested with suitable restriction enzymes and was gel purified. PpCesA6 knockout construct was digested with ApaI and SacI. The PpCesA7 knockout construct was digested with KpnI and SacI. Linear construct DNA (10–15 µg) was used for each transformation.

The detailed procedure of protoplast transformation is described in Chapter 3. For *Pp*CesA6 knockout selection, 50 mg/l^{-1} G418 was used. For PpCesA7 knockout selection, 50 mg/l^{-1} zeocin was used.

Analysis of transformed plants

Two sets of primers were used during PCR. One set annealed at the 5' side of genomic DNA outward of the homology used to make the construct and the 5'side of the selection marker. The other set annealed at the 3' side of the selection marker and the 3'side of the genomic DNA outward of the homology used to make the construct.

For PpCesA6 knockout transformants analysis, forward primer KOA6-5-1 and reverse primer 35SP-1, forward primer 35ST-1 and reverse primer KOA6-3-1 were used for PCR. KOA6-5-1 was derived from the genomic DNA sequence outside of the

homology that was cloned into the 5'MCS of pTN3. 35SP-1 was derived from the sequence of the CaMV 35S promoter. The sequence of KOA6-5-1 is 5'-GCTGCCTTGG TGTCTTAGGTCAAC-3' and the sequence of 35SP-1 is 5'-TCTCAATAGCCCTTTGG TCTTCTG-3. KOA6-3-1 was derived from the genomic DNA sequence outside of the homology that was cloned into the 3'MCS of pTN3. 35ST-1 was derived from the sequence of CaMV 35S polyadenylation signal. The sequence of KOA6-3-1 is 5'-ÇATCATCACCACAGAGGCATCTCC-3' and the sequence of 35ST-1 is 5'-CAGAGCCATGAATAGGTCTATGACCAA-3'.

For PpCesA7 knockout transformants analysis, the forward primer KOA7-5-2 with the reverse primer 35SP-R1 and the forward primer 35ST-F1 with the reverse primer KOA7-3-1 were used for PCR. KOA7-5-2 was derived from the genomic DNA sequence outside of the homology that was cloned between the KpnI and XhoI sites of p35S-zeo. 35SP-R1 was derived from the sequence of the CaMV 35S promoter of p35S-zeo. The sequence of KOA7-5-2 is 5'-TCTTAACCGCAGCCTTGCCTTGTT-3' and the sequence of 35SP-R1 is 5'TGTGAAGCAAGCCTTGAATCGTCC-3'. 35ST-1 was derived from the sequence of the CaMV 35S polyadenylation signal of p35S-zeo. KOA7-3-1 was derived from the genomic DNA sequence outside of the homology that was cloned between the NotI and SacI sites of p35S-zeo. The sequence of 35ST-F1 is 5'-GGACGGAAGGAAGGAGGAAGCCCA-3' and the sequence of KOA7-3-1 is 5'-GGTAAGTCCCTACTCAATGATCAT-3'.

RT-PCR analysis

After PCR analysis, the candidate knockout transformants underwent RT-PCR analysis. Total RNA was extracted by using RNeasy Plant Mini Kit (Qiagen; Valencia CA, USA). 1µg of total RNA was used. The forward primer RT-A6-F2 and the reverse primer RT-A6-7-R1 were used for analysis of PpCesA6 knockout lines. The forward primer RT-A7-F2 and the reverse primer RT-A6-7-R1 were used for analysis of PpCesA7 knockout lines. The sequence of RT-A6-F2 is 5'-TCATCCGACAGGAGAGC G ATGG-3. The sequence of RT-A7-F2 is 5'- GTTATTCGGCAGGAGGGTGATGA-3'. The sequence of RT-A6-7-R1 is 5'- CGCGAT TCG TCCATTATTGGCAGG-3'. The primers actin-F1 5'-CGGAGAGGAAGTACAGTGTGTGGA-3' and actin-R1 5'-ACCAGCCGTTAGAATTGAGCCCAG-3' were used to amplify an actin sequence from *Physcomitrella* (GenBank accession number AW698983) as internal control (Hara *et al.*, 2001).

4.3 Results

4.3.1 Targeted knockout of PpCesA6

1) PpCesA6 knockout construct

The pTN3 vector containing an nptII cassette (kindly provided by RIKEN Bioresource Center, Japan) was used for producing the PpCesA6 construct (Nishiyama *et al.*, 2000). The pTN3 vector was constructed by cloning a modified nptII resistance marker into the pBluescript SK II (+) vector. The nptII cassette was driven by the CaMV

35S promoter and terminated by the CaMV polyadenylation signal. This vector shows the resistance to G418 (Figure 4-2).

A 1058bp fragment (A1) containing the 5' UTR, the first exon and part of the first intron of PpCesA6 genomic DNA was amplified by PCR. An ApaI site was incorporated into the forward primer and a HindIII site was incorporated into the reverse primer. This fragment was cloned into the 5'multiple cloning site of the pTN3 vector between the ApaI and HindIII sites. A fragment (A2) of 1013 bp located at the 3'end of PpCesA6 genomic DNA (starts 113bp after the stop codon) was also amplified by PCR. A BamHI site was incorporated into the forward primer and a SacI site was incorporated into the reverse primer. This fragment was cloned into the 3'multiple cloning site of the pTN3 vector between the BamHI and SacI sites. The resulting construct was the nptII expression cassette flanked by two different PpCesA6 genomic DNA sequences. The orientation of the nptII gene complied with the orientation of PpCesA6. This construct was named KOCesA6 construct (Figure 4-3).

2) Molecular analysis of PpCesA6 knockout lines

The KOCesA6 construct was digested with ApaI and SacI. The double digested linear construct was used in PEG-mediated protoplasts transformation. After gene targeting, except for 389bp of the coding region that includes part of the first intron, the remaining part of the coding region was deleted.

Gene targeting in *Physcomitrella patens* may occur: (1) by replacing the allele with two homologous recombination (HR) events; or, (2) through targeted insertion (TI) in which HR only occurs in one arm of the knockout vector, accompanied by non-

homologous end-joining (NHEJ) by the other arm of the vector (Kamisugi *et al.*, 2006). In order to select transformants that have been generated by two HR events, two different sets of primers are used in PCR based analysis. Primers derived from the 5'side of the genomic region (which was not used in knockout construct) and the 5'side of the selection marker are used to check if the construct have integrated at the 5'end of the targeted locus. Primers derived from the 3'side of the selection marker and the 3' side of the genomic region (which was not used in knockout construct) are used to check if the construct have integrated at the 3'end of the targeted locus.

The forward primer KOA6-5-1 (derived from the genomic DNA sequence located 106bp before where A1 starts), and the reverse primer 35SP-1 (located at the 35S promoter region of the nptII cassette) were used for the first PCR reaction. The forward primer 35ST-1(derived from the 35S polyadenylation signal sequence of the nptII cassette) and the reverse primer KOA6-3-1 (derived from the sequence located 104bp after where A2 ends) were used for the second PCR reaction (Figure 4-3).

More than 100 strains showed G418 resistance. Among these strains, some strains were generated by two HR events occurred at both ends of the targeted locus. Some strains were generated by targeted insertion (TI), in which that DNA integrated at only one end of the targeted locus by HR with one flanking sequence of the vector accompanied by a NHEJ event at the other. Figure 4-4 shows the analysis of 8 knockout strains. When using KOA6-5-1 and 35SP-1 as primers, an expected 1.2kb long PCR product was obtained in strain KOCesA6-1-2, KOCesA6-1-4, KOCesA6-1-5 and KOCesA6-1-6. This PCR product was extracted after gel electrophoresis and submitted

for DNA sequencing. The DNA sequencing result verified that the homologous fragment A1 in the KOCesA6 construct had integrated at the 5'end of the targeted locus. When using KOA6-3-1 and 35ST-1 as primers, an expected 1.2kb-long PCR product was obtained in strain KOCesA6-1-1, KOCesA6-1-3, KOCesA6-1-4, KOCesA6-1-6, KOCesA6-1-7 and KOCesA6-1-257. This PCR product was also extracted after gel electrophoresis and submitted for DNA sequencing. The DNA sequencing result verified that the homologous fragment A2 in the KOCesA6 construct had integrated at the 3'end of the targeted locus. These results suggest that PpCesA6 gene of strain KOCesA6-1-4 and KOCesA6-1-6 was targeted by two HR events and the coding region was replaced by the nptII cassette. These two strains were used for further analysis.

RT-PCR was carried out to check if PpCesA6 was disrupted. The forward primer RT-A6-F2 and the reverse primer RT-A6-7-R1 were used for RT-PCR. RT-A6-F2 was derived from the first exon just after the start codon of PpCesA6, and RT-A6-7-R1 was derived from the sequence spanning the fifth and sixth exons of PpCesA6. No PpCesA6 transcript was detected in the transformants by RT-PCR with these primers (Figure 4-9, lane 3 and 4). The disappearance of the PpCesA6 transcript suggested that this gene was disrupted.

3) Phenotype analysis of PpCesA6 knockout strain

The PpCesA6 knockout line KOCesA6-1-6 was cultured on BCDAT and BCD media and was examined for phenotypic differences. The knockout line did not show any obvious difference from wild type on the colony growth and morphology of protonemata and gametophores (Figures 4-10, 4-11, 4-12). The length of gametophore

was measured from the first leaf at the base to the last leaf at the tip. The tallest gametophores from colonies growing on both BCD and BCDAT media for 5 weeks were measured. The lengths of gametophores growing on BCD and BCDAT media were 3.82±0.46mm and 3.5±0.3mm, respectively (Table 4-1), while in wild type, the numbers were 3.93±0.45mm and 3.51±0.08 mm, respectively. There was almost no difference in the length between the wild type and PpCesA6 knockout line.

The protonemata, axillary hairs and rhizoids were examined under light microscopy. No obvious difference was observed in these structures between the knockout and wild type plants.

The knockout line was also cultured at 16^{o}C under short light cycle to induce gametangia and sporophyte. The knockout line did not show any obvious difference from wild type on gametangia and sporophyte growth.

4.3.2 Targeted knockout of PpCesA7

1) PpCesA7 knockout construct

The vector p35S-Zeo containing a Zeocin expression cassette (kindly provided by RIKEN Bioresource Center, Japan) was used for producing the PpCesA7 knockout construct. The p35S-Zeo vector was constructed by cloning a Zeocin expression cassette into the pBluescript SK II (+) vector. The zeocin expression cassette was driven by the CaMV 35S promoter and terminated by the CaMV 35S polyadenylation signal. This vector shows the resistance to zeocin (Figure 4-5).

A 980bp fragment containing the 5' UTR and part of the first exon of PpCesA7 genomic DNA was amplified by PCR. A KpnI site was incorporated into the 5' primer and an XhoI site was incorporated into the 3'primer. This fragment was cloned into the p35S-Zeo vector between the KpnI and XhoI sites. A 1032bp fragment starting 180bp after the stop codon of PpCesA7 genomic DNA was amplified by PCR. A NotI site was incorporated into the 5' primer and a SacI site was incorporated into the 3'primer. This fragment was cloned into the p35S-Zeo vector between the NotI and SacI sites. The resulting construct was the zeocin expression cassette flanked by two different PpCesA7 genomic DNA sequences. The orientation of the zeocin expression cassette complied with the orientation of PpCesA7. This construct was named KOCesA7 construct (Figure 4-6).

2) Molecular analysis of PpCesA7 knockout lines

The KOCesA7 construct was digested with KpnI and SacI. The double digested linear construct was used for PEG-mediated protoplast transformation. After gene targeting, nearly the full coding region of PpCesA7 (except for the first 60bp) was deleted.

Two sets of PCR reactions were used in the screening of PpCesA7 stable transformants (Figure 4-6). In the first set of PCR, the forward primer KOA7-5-2 (derived from the sequence located 246bp before A1 starts) and the reverse primer 35SP-R1 (derived from the sequence located at the 35S promoter region of Zeocin expression cassette) were used to check if the construct had integrated at the 5'end of the targeted locus. In the second set of PCR, the forward primer 35ST-F1 (derived from the 35S

polyadenylation signal sequence of Zeocin expression cassette) and the reverse primer KOA7-3-1 (derived from the sequence located 234bp after where A2 ends) were used to check if the construct had integrated at the 3'end of the targeted locus.

Figure 4-7 shows the selection of six PpCesA7 disrupted lines. When using KOA7-5-2 and 35SP-R1 as primers, an expected 1.3kb PCR product was obtained in strain KOCesA7-1 and KOCesA7-3. This PCR product was extracted after gel electrophoresis and submitted for DNA sequencing. The DNA sequencing result verified that the fragment A1 in KOCesA7 construct had integrated at the 5'end of the targeted locus. When using KOA7-3-1 and 35ST-F1 as primers, an expected 1.7kb PCR product was obtained in strain KOCesA7-3. The PCR product was also extracted after gel electrophoresis and submitted for DNA sequencing. The DNA sequencing result verified that the fragment A2 in KOCesA7construct had integrated at the 3'end of the targeted locus. These results suggested that PpCesA7 gene of strain KOCesA7-3 was targeted by two HR events and the coding region was replaced by the zeocin expression cassette. The KOCesA7-3 strain was used for further analysis.

RT-PCR was carried out to check if PpCesA7 was disrupted. The forward primer RT-A7-F2 and the reverse primer RT-A6-7-R1 were used for RT-PCR. RT-A7-F2 was derived from the first exon just after the start codon of PpCesA7 and RT-A6-7-R1 was derived from the sequence spanning the fifth and sixth exons of PpCesA7. No PpCesA7 transcript was detected in KOCesA7-3 by RT-PCR with these primers (Figure 4-9, lane 5 and 7). The disappearance of the PpCesA7 transcript confirmed that PpCesA7 was disrupted.

3) Phenotype analysis of PpCesA7 knockout strain

The KOCesA7-3 knockout line was cultured on BCD and BCDAT media and was examined for phenotypic differences. As in the PpCesA6 knockout line, the PpCesA7 knockout line did not show any obvious difference from wild type on the colony growth and morphology of protonemata and gametophores (Figures 4-10, 4-11, 4-12). The length of the gametophore was measured as well. The tallest gametophores from colonies growing on both BCD and BCDAT media for 5 weeks were measured. The lengths of the knockout line cultured on BCD and BCDAT media were 3.65±0.68mm and 3.73±0.49mm, respectively, while in the wild type, the numbers were 3.93±0.45mm and 3.51±0.08mm, respectively (Table 4-1), so there was no obvious difference of the length between the wild type and the PpCesA7 knockout line.

The protonemata, axillary hairs and rhizoids were examined under light microscopy. No obvious differences were observed in these structures between the knockout and the wild type plants.

The knockout lines were also cultured at 16^{o}C under short light cycle to induce gametangia and sporophyte. The knockout lines did not show any obvious difference from wild type on gametangia and sporophyte growth.

4.3.3 PpCesA6 and PpCesA7 double knockout

1) Molecular analysis of PpCesA6/PpCesA7 double knockout lines

Single gene knockout of both PpCesA6 and PpCesA7 did not show obvious phenotype. Double gene knockout was performed to further investigate the function of

PpCesA6 and PpCesA7. The PpCesA6 knockout strain KOCesA6-1-6 was used for generating double knockout strains. The double restriction enzyme digested linear KOCesA7 construct was introduced into the PpCesA6 knockout line KOCesA6-1-6 protoplast by PEG-mediated transformation.

Primers KOA7-5-2 and 35SP-R1were used to ascertain if the fragment A1 in the KOCesA7 construct had integrated at the 5'end of the targeted locus. Primers KOA7-3-1 and 35ST-F1 were used to check if the fragment A2 in the KOCesA7 construct had integrated at the 3'end of the targeted locus. Figure 4-8 shows the analysis of 10 double knockout transformants. Knockout strains DD-3 and DD-6 were generated by two HR events.

RT-PCR was carried out to confirm that PpCesA7 was disrupted. The forward primer RT-A7-F2 and the reverse primer RT-A6-7-R1 were used in RT-PCR. No PpCesA7 transcript was detected in the transformant. The disappearance of the PpCesA7 transcript confirmed that PpCesA7 was disrupted. Another RT-PCR was performed with the forward primer RT-A6-F2 and the reverse primer RT-A6-7-R1 to confirm that PpCesA6 was indeed disrupted. No PpCesA6 transcript was detected (Figure 4-9, lane 8 and 9). The disappearance of both PpCesA7 and PpCesA6 transcript confirmed that both genes were disrupted in the same transformant.

2) Phenotype analysis of PpCesA6/PpCesA7 double knockout strain

The double knockout line was cultured on BCD and BCDAT media and was examined for phenotypic difference. The knockout lines did not show any obvious difference from wild type on the colony growth and morphology of protonemata and

gametophores (Figures 4-10, 4-11, 4-12). However, a closer exam revealed that the colonies of the double knockout mutant appeared to be more compact than wild type and single knockout lines. The tallest gametophores from colonies growing on both BCD and BCDAT media for 5 weeks were measured. The lengths of double knockout mutants growing on BCD and BCDAT medium were 2.75±0.22mm and 2.65±0.43mm, respectively. The lengths of the double knockout mutants were significantly shorter than the wild type as well as the single knockout mutant (Table 4-1).

The protonemata, axillary hairs and rhizoids were examined under light microscopy. No obvious differences were observed in these structures.

The knockout lines were also cultured at 16^{o}C under short light cycle to induce gametangia and sporophyte. The knockout lines did not show any obvious difference from wild type on gametangia and sporophyte growth.

4.4 Discussion

Through this study, the PpCesA6 and PpCesA7 single knockout line as well as the double knockout line were generated. A phenotypic analysis showed that there was no obvious difference on the colony growth and morphology of gametophores. A previous study showed that PpCesA6 is expressed in protonemata, axillary hairs and rhizoids. The light microscopy and scanning electron microscopy studies did not reveal obvious difference in these structures. A closer exam showed that the double knockout mutant colonies were more compact than wild type and single knockout lines. Further study showed that the compact appearance was caused by reduction in the stem length. In

single knockout lines of both genes, gametophores growing on both BCD and BCDAT media did not show any reduction of the stem length, while in double knockout line, the reduction of stem length was very obvious. Compared with wild type, the reductions for double knock mutant growing on BCD and BCDAT media were 30% and 24%, respectively.

Among *Arabidopsis* primary cell wall biosynthesis genes *At*CesA1, 2, 3, 5, 6 , *At*CesA5 and *At*CesA2 are partially redundant with *At*CesA6 and they most likely compete with *At*CesA6 for the same binding site in the cellulose synthase complex (Burn *et al.*, 2002; Desprez *et al.*, 2002, 2007; Persson *et al.*, 2007; Somerville, 2006). *At*CesA4, *At* CesA7 and *At*CesA8 are involved in the secondary wall synthesis (Taylor *et al.*, 2000, 2003). So far there has been no report on gene function redundancy of secondary wall biosynthesis.

PpCesA6 and PpCesA7 are highly similar to each other. They encode almost identical protein. It is very likely that they have the same function. This study showed that PpCesA6 and PpCesA7 are redundant. When one gene is mutated, the other one compensates for the function. No phenotypic difference is observed. When both genes are mutated, the mutant shows obvious phenotypic differences from the wild type. The double knockout mutants did not show any severe phenotype. Given that there are 9 functional CesAs in *Physcomitrella*, it is likely that PpCesA6 and PpCesA7 may be partially redundant with other CesAs.

4.5 Acknowledgements

I would like to thank Dr. Tomaki Nishiyama and Dr. Yuji Hiwatashi at the National Institute for Basic Biology, Myoudaiji-cho, Okazaki-shi, Japan for giving me permission to use the pTN3 vector and the p35S-Zeo vector, respectively. I would also like to thank RIKEN Bioresource Center, Tsukuba-shi, Ibaraki Japan for providing me these vectors.

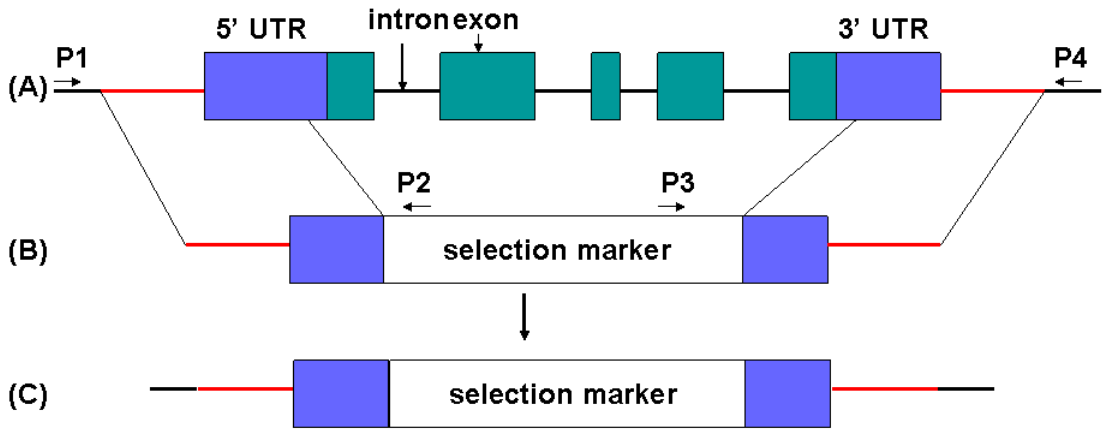


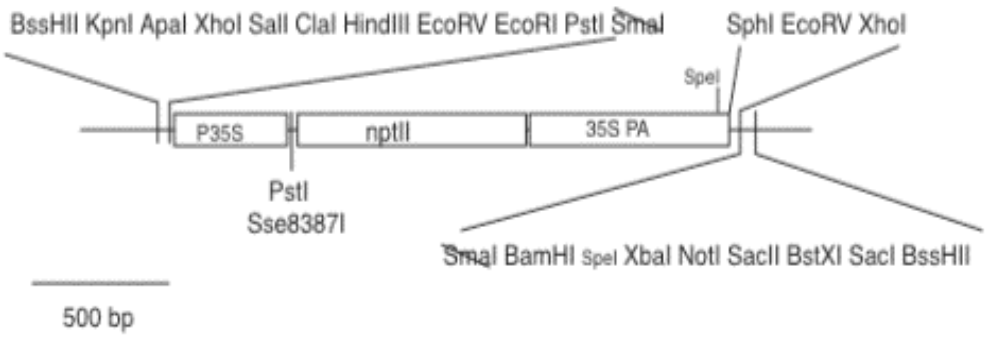
Figure 4-1 Strategy of targeted knockout

(D) Structure of the gene of interest. Exons are in dark green rectangles. UTRs are in blue rectangles. Lines between squares indicate introns.

(E) Structure of knockout construct. On the left side of the selection marker is the first homologous fragment. On the right side of the selection marker is the second homologous fragment.

(F) The outcome after homologous recombination.

Two sets of primers are used for PCR based analysis. Primers P1 and P2, Primers P3 and P4 are used together, respectively.



P35S: CaMV 35S promoter

35S PA: CaMV polyadenylation signal

Figure 4-2 pTN3 vector

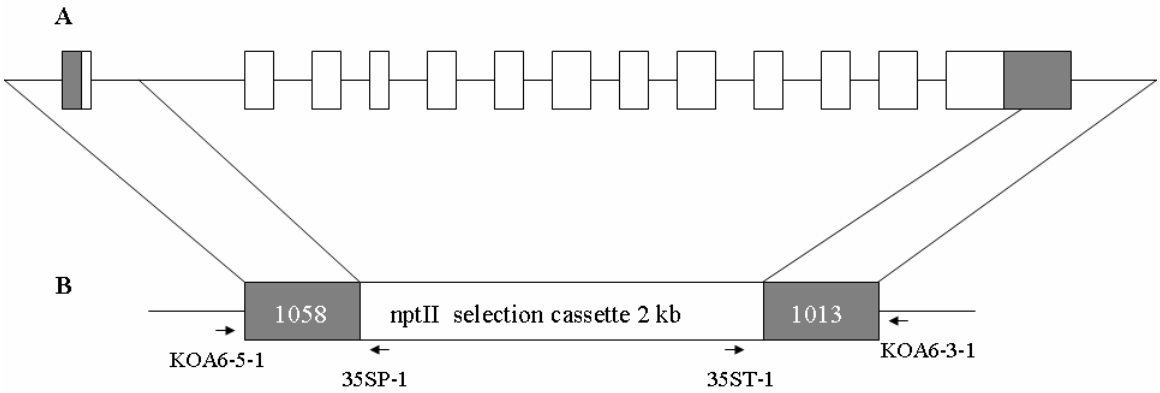


Figure 4-3 Schematic representation of the PpCesA6 gene and KOCesA6 knockout construct

(G) Structure of PpCesA6 gene from *P. Patens.* The white rectangles represent exons, grey rectangles represent UTR, lines represent introns.

(H) The KOCesA6 knockout construct. The 1058bp fragment was cloned between the ApaI and HindIII sites of pTN3. The 1013bp fragment was cloned between the BamHI and SacI sites of pTNs.The positions of the PCR primers are indicated by arrows.

Primers KOA6-5-1 and 35SP-1, 35ST-1 and KOA6-3-1 were used together, respectively.

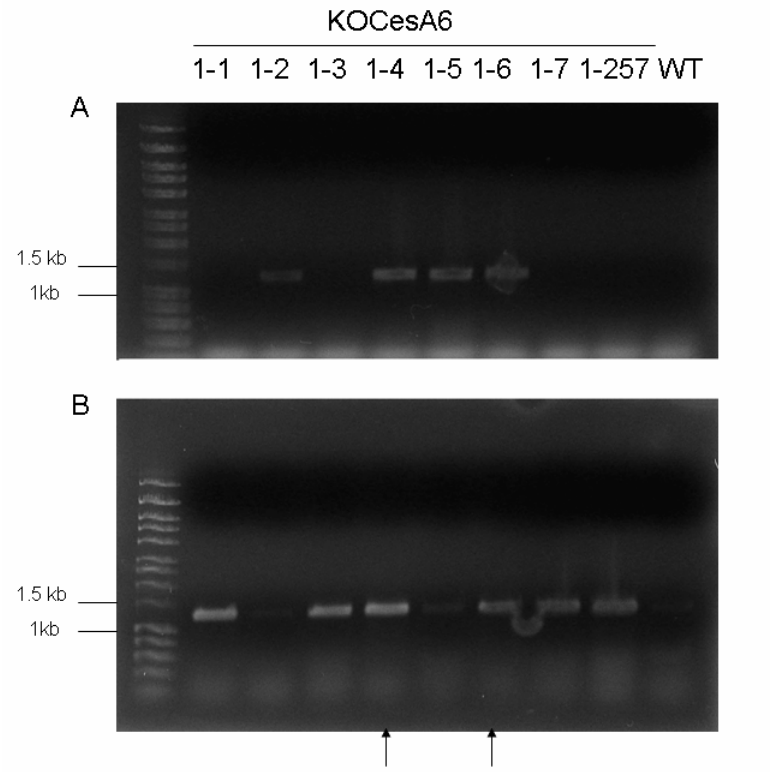


Figure 4-4 PCR analysis of PpCesA6 knockout strains

A PCR product from the 5' side genomic region primer KOCesA6-5-1 and the 5' side selection cassette primer 35SP-1

B PCR product from the 3' side selection cassette primer 35ST-1 and the 3' genomic region primer KOCesA7-3-1

Note that the knockout strains KOCesA6-1-4 and KOCesA6-1-6 have both PCR products.

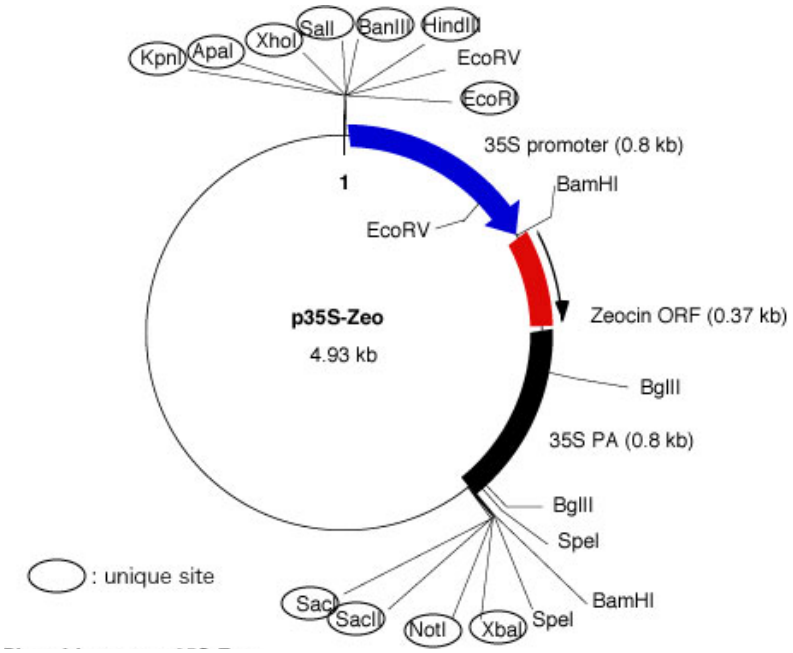


Figure 4-5 p35S-Zeo vector

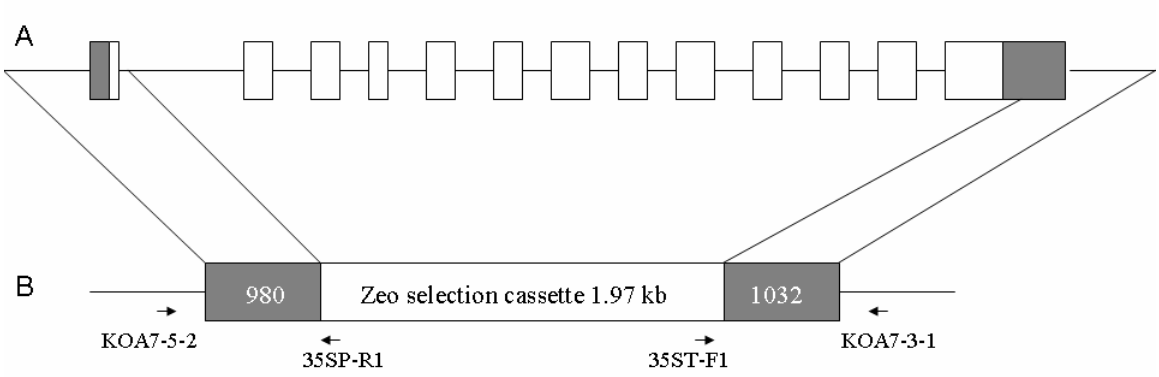


Figure 4-6 Schematic representation of the PpCesA7 gene and KOCesA7 knockout construct

(I) Structure of PpCesA7 gene from *P. Patens*. The white rectangles represent exons, grey rectangles represent UTR, lines represent introns.

(J) The KOCesA7 knockout construct. The 980bp fragment was cloned between the KpnI and XhoI sites of p35S-Zeo. The 1032bp fragment was cloned between the NotI and SacI sites of p35S-Zeo. The positions of the PCR primers are indicated by arrows.

Primers KOA7-5-2 and 35SP-R1, 35ST-F1 and KOA7-3-1 were used together, respectively.

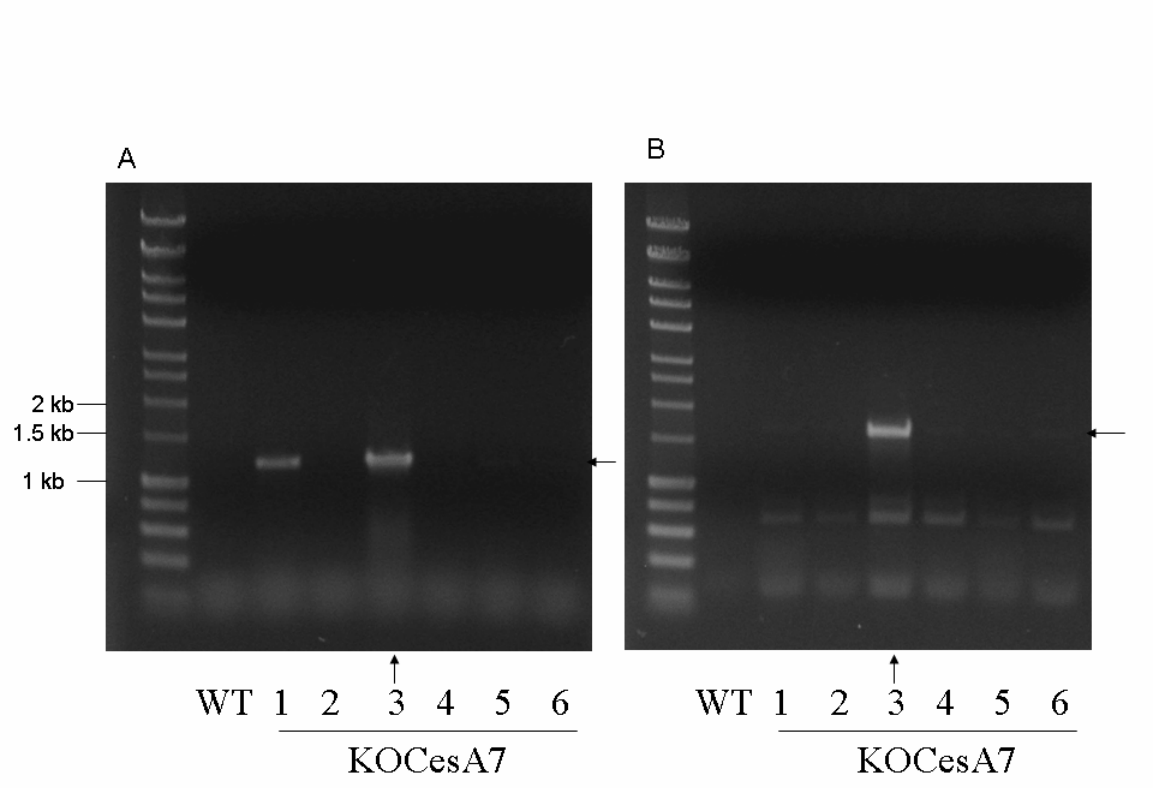


Figure 4-7 PCR analysis of PpCesA7 knockout strains
A PCR product from the 5' side genomic region primer KOCesA7-5-2 and the 5' side selection cassette primer 35SP-R1. Arrow indicates the 1.3kb PCR product.
B PCR product from the 3' side selection cassette primer 35ST-F1 and the 3' genomic region primer KOCesA7-3-1. Arrow indicates the 1.7kb PCR product.
Note that the knockout strain KOCesA7-3 has both PCR products.

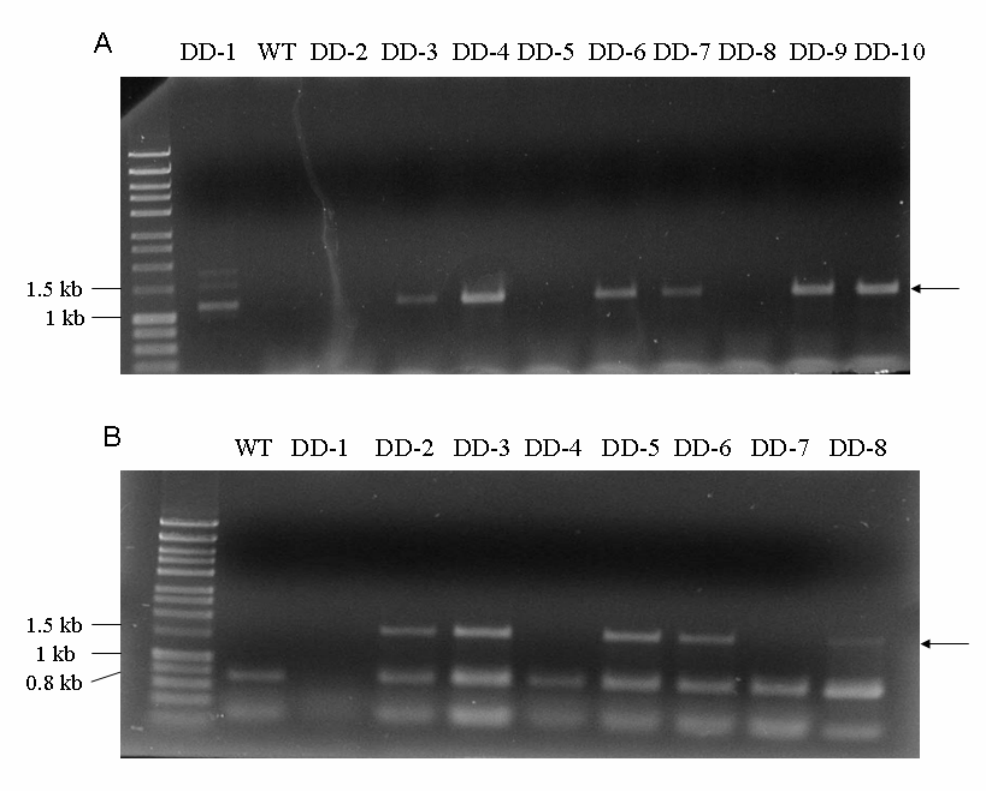


Figure 4-8 PCR analysis of double knockout strains

A Primers KOA7-5-2 and 35SP-R1 were used to screen PpCesA6 knockout line transferred with KOCesA7 construct. Note the 1.3 kb PCR product.

B Primers 35ST-F1 and KOA7-3-1 were used to screen PpCesA6 knockout line transferred with KOCesA7 construct. Note the 1.7 kb PCR product. The 0.7kb product was caused by mispriming. Knockout lines DD-3 and DD-6 were generated by two HR events.

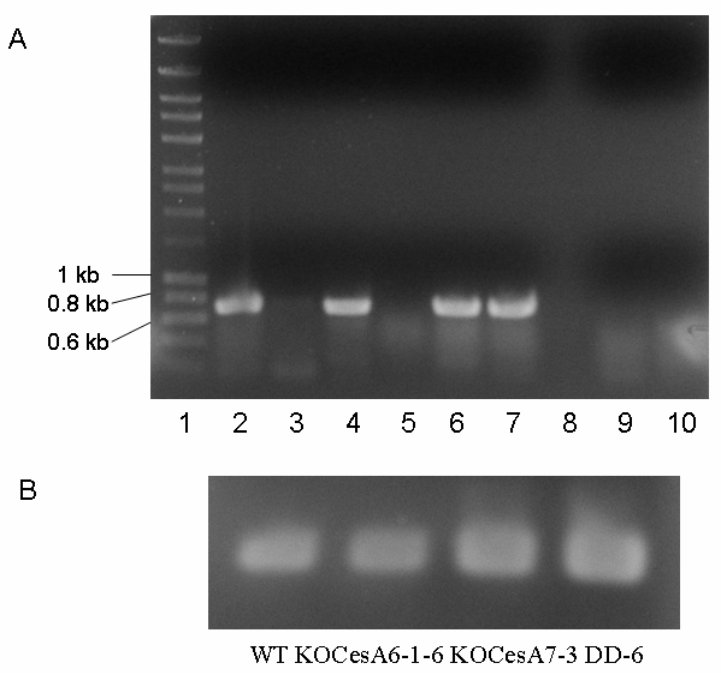
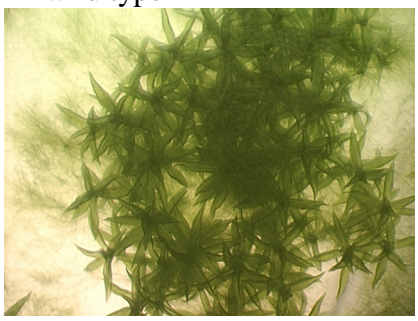


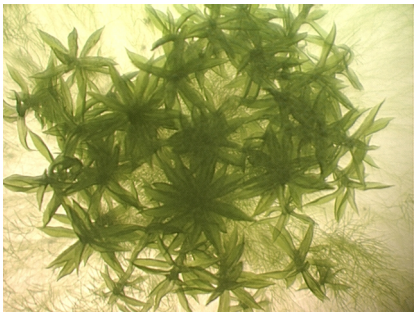
Figure 4-9 RT-PCR analysis of knockout lines

A lane 1 DNA marker
Lane2 wild type with RT-A6-F2 and RT-A6-7-R1 as primers
Lane 3 KOCesA6-1-6 knockout strain with RT-A6-F2 and RT-A6-7-R1 as primers
Lane 4 KOCesA6-1-6 knockout strain with RT-A7-F2 and RT-A6-7-R1 as primers
Lane 5 KOCesA7-3 knockout strain with RT-A7-F2 and RT-A6-7-R1 as primers
Lane 6 wild type with RT-A7-F2 and RT-A6-7-R1 as primers
Lane 7 KOCesA7-3 knockout strain with RT-A6-F2 and RT-A6-7-R1 as primers
Lane 8 Double knockout strain DD-6 with RT-A7-F2 and RT-A6-7-R1 as primers
Lane 9 Double knockout strain DD-6 with RT-A6-F2 and RT-A6-7-R1 as primers
Lane 10 blank
B actin gene from *Physcomitrella* as control

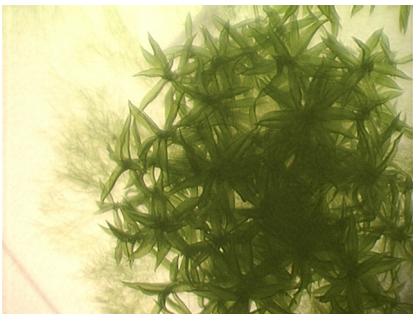
A wild type



B KOCesA6-1-6



C KOCesA7-3



D double knockout DD-6

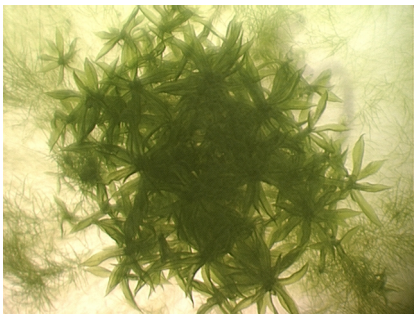
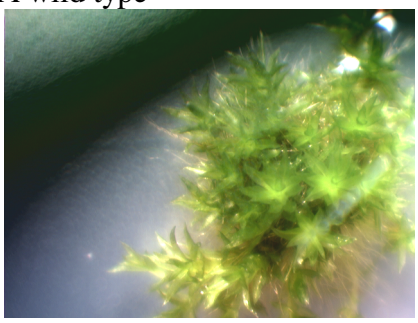
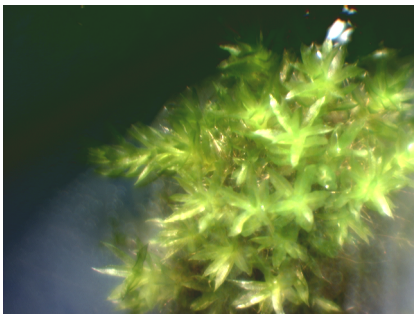


Figure 4-10 Moss colonies growing on BCDAT medium

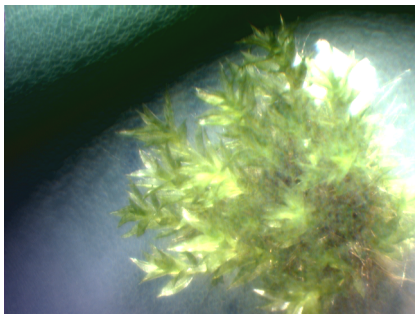
A wild type



B KOCesA6-1-6



C KOCesA7-3



D double knockout DD-6

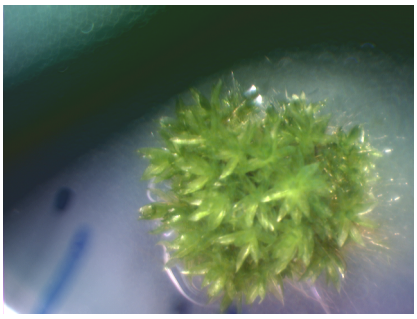


Figure 4-11 Moss colonies growing on BCD medium

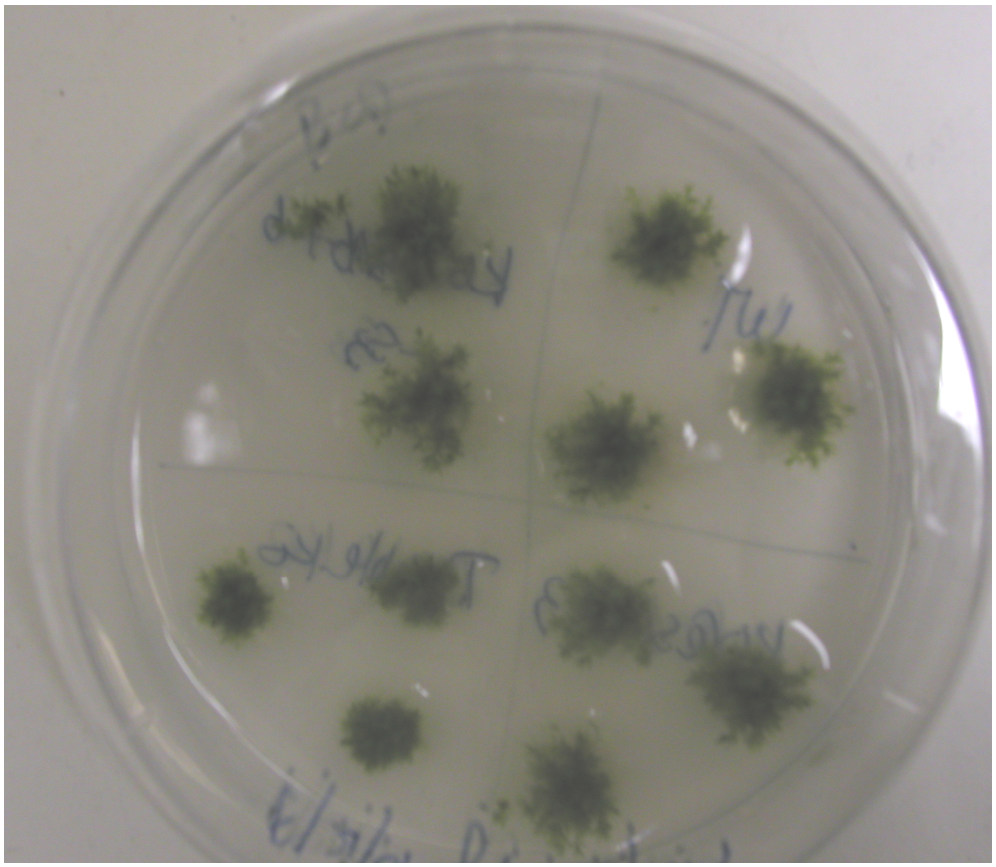


Figure 4-12 Comparison of moss colonies growing on BCD medium
The quadrangles are listed clockwise from the top right corner: WT, PpCesA7 knockout line KOCesA7-3, double knockout line DD-6, PpCesA6 knockout line KOCesA6-1-6. Note that the colonies of the double knockout line (lower left quadrangle) are more compact.

Table 4-1 Length of gametophores growing on BCD and BCDAT media

	WT	KOCesA6	KOCesA7	Double knockout
BCD	3.93±0.45mm	3.82±0.46mm	3.65±0.68mm	2.75±0.22mm
BCDAT	3.51±0.08mm	3.5±0.3mm	3.73±0.49mm	2.65±0.43mm

WT-wild type
KOCesA6-PpCesA6 knockout line
KOCesA7-PpCesA7 knockout line
Double knock-PpCesA6/PpCesA7 knockout line

Chapter 5 Identification and characterization of cellulose synthase-like C4 (CslC) gene in *Physcomitrella patens*

5.1 Introduction

Cellulose synthase-like (CSL) genes encode proteins that show structural similarity to cellulose synthase. CSL proteins are also integral membrane proteins and they also contain the D, D, D, QXXRW motif (Cutler and Somerville, 1997; Saxena and Brown, 1997). Together with cellulose synthase genes, they belong to the same superfamily. There are 8 CSL subfamilies, CslA to CslH (Hazen *et al*, 2002; Richmond and Somerville, 2000). Compared with CesA proteins, CSL proteins are smaller proteins (except CslDs). They have less than 1000 amino acids. Unlike CesA proteins, CSL proteins have no plant conserved region CP-R, nor do they have the Zn-binding domain in the N-terminal region. It is believed that CSL proteins are involved in the synthesis of non-cellulosic polysaccharides (Richmond and Somerville, 2001).

When this study started, the function of CslC genes was unknown. Before this study was completed, Cocuron *et al.* proved that CslC proteins from nasturtium and *Arabidopsis* have glucan synthase activity. CslC proteins produce β-1, 4-linked glucans that are involved in the synthesis of the xyloglucan (XyG) backbone (Cocuron *et al.*, 2007).

The goals of this study were to: (1) identify and characterize this group of genes in *Physcomitrella*; and (2) perform targeted gene knockout to study the function of some of these genes.

5.2 Materials and methods

EST and genomic database search

Arabidopsis CslC sequences were used as queries to search against PHYSCObase (http://moss.nibb.ac.jp). PHYSCObase was the first available public database containing *Physcomitrella* EST sequences generated from the regenerating protoplasts library, the untreated protonemata library, the auxin-treated library, the cytokinin-treated library and the developing sporophytes library. Contigs with the e-value above 1.0 were undergone further analysis. These contigs were submitted to GenBank for BLASTX search. The sequences identified as CslCs were the most interested ones. Potential full-length clones were selected and these clones were obtained from RIKEN BioResource Center, Tsukuba-shi, Ibaraki, Japan. The EST clones were sequenced from both strands and the cDNA sequences were obtained.

Subsequently, the CslC cDNA sequences were used in megablast (http://www.ncbi.nlm.nih.gov/) search against *Physcomitrella* genome sequences in GenBank that were released from the whole genome shotgun project carried by the Joint Genome Institute of U.S. Department of Energy while sequencing.

All the searches were repeated with the releases of more EST sequences and raw genomic sequences.

PpCslC4 knockout construct

The pTN3 vector was used for producing PpCslC4 knockout construct. The forward primer Kcslc-A1-F1-KpnI with a KpnI site and a reverse primer Kcslc-A1-R1-ApaI with an ApaI site were used to amplify a fragment containing the 5'-UTR and the first exon. The sequence of Kcslc-A1-F1-KpnI is 5'-GCCGGTACCTAATCTTCATGCATGGAAGGC-3'. The sequence of Kcslc-A1-R1-ApaI is 5'-AATGGGCCCGCTCGTTGCACATGGGTATTT-3'. The sequences of the added restriction sites were underlined. This fragment was cloned into the 5'MCS of pTN3. The forward primer Kcslc-A2-F1-BamHI with a BamHI site and the reverse primer Kcslc-A2-R1-NotI with a NotI site were used to amplify a second fragment containing the 3'-UTR of PpCslC4. The sequence of Kcslc-A2-F1-BamHI is 5'-GCCGGATCCTCAGGTATTACACGAAAGTAG-3'. The sequence of Kcslc-A2-R1-NotI is 5'-TAAGCGGCCGCCACTTTCATCTTGATACGG-3'. The sequences of the added restriction sites were underlined. This fragment was cloned into the 3'MCS of pTN3. DNA sequencing was carried out to confirm that there were errors during amplifying and cloning the sequences.

***Physcomitrella* protoplast transformation**

The knockout construct was digested with KpnI and NotI and was gel purified. Linear construct DNA (10–15 µg) was used for each transformation. The detailed procedure of protoplast transformation is described in Chapter 3. 50 mg/l^{-1} G418 was used in selection.

Molecular analysis of PpCslC4 knockout lines

Two sets of primers were used during PCR. One set annealed at the 5' side of genomic DNA outward of the homology used to make the construct and the 5'side of the selection marker. The other set annealed at the 3' side of the selection marker and the 3'side of the genomic DNA outward of the homology used to make the construct.

In the first set of PCR, a forward primer KOCslC-5-1 (derived from the genomic DNA sequence outside of the homology that was cloned into the 5'MCS of pTN3) and a reverse primer sp-35SP-2 (derived from the sequence located at the 35S promoter region of nptII cassette) were used to check if the construct had integrated at the 5'end of the targeted locus. In the second set of PCR, a forward primer sp-35ST-2 (derived from the 35S polyadenylation signal sequence of nptII cassette) and a reverse primer KOCslC-3-1 (derived from the genomic DNA sequence outside of the homology that was cloned into the 5'MCS of pTN3) were used to check if the construct had integrated at the 3'end of the targeted locus. Primer sequences were as follows:

KOCslC-5-1 5'-GCTGTAGGACAGATTCCAACATTG-3',

KOCslC-3-1 5'-ATCCTTAATGACTGACAATAGCAC-3',

sp-35SP-2 5'-CGAGGAGGTTTCCCGAAATTACCC-3',

sp-35ST-2 5'-GGTATCAGAGCCATGAATAGGTCT-3'.

RT-PCR analysis

Total RNA was extracted by using a RNeasy Plant Mini Kit (Qiagen, Valencia CA, USA). 1µg of total RNA was used. The forward primer RT-CslC-F1 and the reverse primer RT-CslC-R1 were used for analysis of PpCslC4 knockout lines. RT-

CslC-F1 was derived from the sequence just after the start codon. RT-CslC-R1 was derived from the sequence of the second exon. The sequences of RT-CslC-F1 and RT-CslC-R1 are 5'-AGGATGCCGTCGAAGCTTGAATTT-3'and 5'-CCTGGATCAAGAT CCGGTTCTTCG –3', respectively. The primers actin-F1 5'-CGGAGAGGAAGTACAG TGTGTGGA-3' and actin-R1 5'-ACCAGCCGTTAGAATTGAGCCCAG-3' were used to amplify an actin sequence from *Physcomitrella* (GenBank accession number AW698983) as internal control (Hara *et al.*, 2001).

5.3 Results

5.3.1 PpCslC4 gene

When using *Arabidopsis* CslC genes as queries searching against the PHYSCObase, a contig corresponding to clone pphn1p17 was one of several contigs that had the strongest hits. Further sequence analysis showed that this clone had both the 5'-UTR and 3'-UTR. It was a potential full-length cDNA clone. This clone was obtained from RIKEN BioResource Center (Japan). Both strands of this clone were sequenced, and the full-length sequence was obtained. This full- length cDNA sequence is different from the other three *Pyhscomitrella* CslCs that have been deposited in GenBank (Roberts and Bushoven, 2007). This sequence was named PpCslC4. The genomic DNA sequence of PpCslC4 was obtained from *Physcomitrella* genome database (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html) by using the cDNA sequence as the query in searching this database. Both the cDNA sequence and genomic DNA sequence of

PpCslC4 were deposited in GenBank. The accession numbers for the cDNA sequence and the genomic DNA sequence are EF608234 and EF608235, respectively.

PpCslC4 cDNA is 3389 bp long with a coding region of 2085 bp. The ORF starts from ATG at position 778 and stops at position 2862. The 5'-UTR contains 777 bp and the 3'-UTR contains 527 bp. The protein it encodes is 694 amino acids in length. The nucleotide and derived amino acid sequences of PpCslC4 is shown in Figure 5-1. The D, D, D, QXXRW motif is in the form of DDS, DAD, VED, QQHRW. The Zinc-finger domain is not found near the N-terminus.

The PpCslC4 genomic region is 4196 bp in length. It has 4 small introns in the coding region. There is also one small intron in the 5'-UTR. The gene structure of PpCslC4 is shown in Figure 5-2.

The deduced amino acid sequence of PpCslC4 shows 72% similarity with PpCslC2 and PpCslC3 (GenBank accession number DQ898285 and DQ898286), and 65% similarity with PpCslC1 (GenBank accession number DQ898284). When compared with other organisms, PpCslC4 protein shows more than 60% similarity with *Arabidopsis* and *Oryza sativa* CslC proteins. The multiple alignments of the deduced amino acid sequences of PpCslC4 with other *Physcomitrella* CslCs and *Arabidopsis* and *Oryza sativa* CslCs by CLUSTALX (Thompson *et al.*, 1997) is shown in Figure 5-3.

5.3.2 PpCslC4 knockout construct

The vector pTN3 containing an nptII cassette was used for producing PpCslC4 knockout construct (Nishiyama *et al.*, 2000). A 947bp fragment (A1) of PpCslC4

genomic DNA (including 5'UTR and the first exon) was amplified by PCR with primers KcslcA1-F1-KpnI and Kcslc-A1-R1-ApaI. This fragment was cloned into the 5'multiple cloning sites of pTN3 between the ApaI and KpnI sites. A fragment (A2) of 965 bp located at the 3'end just 2bp downstream of the stop codon of PpCslC4 genomic DNA was amplified by PCR with primers Kcslc-A2-F1-BamHI and Kcslc-A2-R2-NotI. This fragment was cloned into the 3'multiple cloning sites of pTN3 between the BamHI and NotI sites. The resulting construct with the nptII expression cassette was flanked by two different PpCslC4 genomic DNA sequences. The orientation of the nptII gene complied with the orientation of PpCslC4. The construct was named KOCslC4 construct (Figure 5-5). This construct is a replacement construct.

5.3.3 Molecular analysis of PpCslC4 knockout lines

The KOCslC4 construct was digested with KpnI and NotI. The double restriction enzyme digested linear construct was used for the PEG-mediated protoplast transformation. Two sets of PCR reactions were used in the screening of stable transformants. In the first set of PCR, a forward primer KOCslC5-1 derived from the sequence located 311bp upstream of A1 and a reverse primer sp-35SP-2 derived from the sequence located at the 35S promoter region of nptII cassette were used to check if the construct had integrated at the 5'end of the targeted locus. In the second set of PCR, a forward primer sp-35ST-2 derived from the 35S polyadenylation signal sequence of nptII cassette and a reverse primer KOCslC-3-1 derived from the sequence located 243bp

downstream of A2 were used to check if the construct had integrated at the 3'end of the targeted locus.

Figure 5-6 shows the selection of ten PpCslC4 knockout lines (KOCslCDD- 1-10). When using primers KOCslC-5-1 and sp-35SP-2, an expected 1.3kb-long PCR product was obtained in all the lines. This product was extracted after gel electrophoresis and was submitted for DNA sequencing. The DNA sequencing result verified that the homologous fragment A1 in KOCslC4 had integrated into the 5' side of the targeted locus. However, when using KOCslC-3-1 and sp-35ST-2 as primers, no PCR product was obtained. Gene targeting in *Physcomitrella patens* may occur: (1) by replacing the allele with two homologous recombination (HR) events or (2) through targeted insertion (TI) in which HR only occurs in one arm of the knockout vector, accompanied by non-homologous end-joining (NHEJ) by the other arm of the vector (Kamisugi *et al.*, 2006). This result suggests that targeted insertion has occurred in generating all of the 10 PpCslC4 knockout lines that are shown in Figure 5-6. Another transformant line KOCslCDD-11 did not show any PCR product when using either primers KOCslC-5-1 and sp-35SP-2 or primers KOCslC-3-1 and sp-35ST-2. This result suggests PpCslC4 in KOCslCDD-11 transformant line was not disrupted.

In order to verify that PpCslC4 was disrupted, RT-PCR analysis was carried out. Knockout lines KOCslCDD-2, KOCslCDD-4 and KOCslCDD-11 were used in RT-PCR analysis. Primers derived from PpCslC4 exon sequences were used in RT-PCR. RT-PCR products were obtained in wild type strain as well as in KOCslCDD-11 strain. No RT-PCR product was obtained in KOCslCDD-2 and KOCslCDD-4 strains (Figure 5-7).

The disappearance of PpCslC4 transcript in these two strains confirmed that PpCslC4 was disrupted.

5.3.4 Phenotype analysis of knockout lines

The PpCslC4 knockout strains were cultured on BCDAT and BCD media for phenotypic analysis. The knockout mutants did not show obvious difference from the wild type for colony growth and morphology of protonemata and gametophores.

5.4 Discussion

In this study, a new CslC gene in *Physcomitrella* was identified. Together with the other CslC genes that have been identified, there are 4 identified CslC genes in this organism (Roberts and Bushoven, 2007). Five CslC genes have been identified in *Arabidopsis* (Hamann *et al.*, 2004). There are almost as many CslC genes in *Physcomitrella* as in *Arabidopsis*. The study of CslC genes is very limited. There are not many CslC gene sequences available in GenBank. Among all the CSL families, CslCs are more similar to bacterial cellulose synthase genes than to plant cellulose synthase genes (Nobles and Brown, 2004).

The plant cell walls contain cellulose, hemicellulose, pectin and lignin. Cellulose is the major component of the cell wall. Hemicellulose includes xyloglucan (XyG), glucomannan, glucuronoarabinoxylan (GAX) and mixed linked glucan (MLG) (Carpita and McCann, 2000). Cellulose synthase (CesA) proteins are involved in the biosynthesis of cellulose. Cellulose synthase-like (CSL) proteins are thought to be involved in the

biosynthesis of non-cellulosic polysaccharides. The function of CslC genes has been discovered recently. Developing nasturtium (*Tropaeoleum majus*) seeds accumulate a large amount of xyloglucan (XyG). By screening a developing nasturtium seed cDNA library, a CslC gene was found to be overexpressed. Expression of this gene and AtCslC4 (its closest homolog in *Arabidopsis*) in yeast resulted in the synthesis of β-1, 4-glucans. This result demonstrated that CslC proteins encode XyG glucan synthases (Cocuron *et al.*, 2007).

The machinery (the rosette terminal complex) for making cellulose is complicated. More than one gene is required in cellulose biosynthesis. At least three cellulose synthase proteins are playing distinct roles within the rosette (Taylor *et al.*, 2000, 2003). Compared with cellulose biosynthesis, the biosynthesis of other cell wall polysaccharides is simpler. One gene can be involved in different tasks. For example, recent studies showed that CslA proteins are involved in both mannan and glucomannan biosynthesis (Liepman, 2005; Liepman *et al.*, 2007). It is very likely that other CSL proteins are involved in more than one task.

By taking advantage of homologous recombination in *Physcomitrella*, PpCslC4 gene was disrupted. No obvious phenotypic differences were observed in the knockout plants. Given that there are 4 CslC genes in *Physcomitrella* and CslA proteins can play two different roles in cell wall polysaccharide synthesis, it is very likely that other CslC genes or other Csl genes compensate for the function.

5.5 Acknowledgements

I would like to Dr. Tomaki Nishiyama at the National Institute for Basic Biology, Myoudaiji-cho, Okazaki-shi, Japan for giving me permission to use the pTN3 vector and RIKEN Bioresource Center, Tsukuba-shi, Ibaraki, Japan for providing me EST clones.

```
agcggccctgtgatgatgtttttggagttttgccactgttttctggagagtgtttcgagt 60
attgcgaattttgggcggggttttcgccatttggggctttttgatgtgaatgtcaggcgg 120
tggcgattatcagccacgaccggaacgttcaggctttggacggccctccctgtattcata 180
ccgagttttgcgtggagtccctgcacgggatgacgcgctaatgctggggagggagatcac 240
actttcccctggggactgaatgagtggtgtcgcagggtgtcttggaatgtgagcgattgg 300
tttaggagccactgcagggtttctggttgtgtattatgtctgaagcatatgccacggctt 360
ccctctgctttctctttctgcgctaccatcaccacagctcgaaagctgtaggacagattc 420
caacattgttcacttcgttgttcacagatttagaatgctggtcggttgttgaggttttcc 480
tttagggacacattgagcagttctttttcaagttttcagtggtgtacgtgcctaatcttc 540
atgcatggaaggcaccgagagattgcaaccgtttctgttacatagaagagtgctgtttac 600
gtaggtggctttaaaggggcgacagaagtacctcaaacttgctaccatttgtgactcct 660
gatggattacatttcccgttaactatcatcagtccgcagcattggctatagtaggcctta 720
gctttagccgaccgtcaacaagttgacaactgcaatctacaccttcatcctcacagg atg 780
                                                          M  1
ccgtcgaagcttgaattttccgaattatttggtaaggagtcaaatcggggaccaccagtg 840
 P  S  K  L  E  F  S  E  L  F  G  K  E  S  N  R  G  P  P  V  21
gtggtgaagatggagaatcccaacttcgatatgctggagtttgatagtcccacgactgcg 900
 V  V  K  M  E  N  P  N  F  D  M  L  E  F  D  S  P  T  T  A  41
atgcgtggtggacaggcagccaaaggcaagggcaaaaatgcgaagcagttaacctgggtg 960
 M  R  G  G  Q  A  A  K  G  K  G  K  N  A  K  Q  L  T  W  V  61
ctgcttctcaaagccaacaaggctgttggatgcttgacgtggttagcatctggtattatg 1020
 L  L  L  K  A  N  K  A  V  G  C  L  T  W  L  A  S  G  I  M  81
atattgttggacgccatcaaggatcgactcattctccgcaaaaatgtcatcaacgcttcg 1080
 I  L  L  D  A  I  K  D  R  L  I  L  R  K  N  V  I  N  A  S  101
aagggaaagttatcccgggttatcatcggatttctgatctttgcgttgattatgttgtgc 1140
 K  G  K  L  S  R  V  I  I  G  F  L  I  F  A  L  I  M  L  C  121
gttgaagtgggagctcacaccctaggctggcaattctccgtgccacaatggccaaccact 1200
 V  E  V  G  A  H  T  L  G  W  Q  F  S  V  P  Q  W  P  T  T  141
cttaatgtttcaagtcttccgcatgcgttattcatgggatggttgtttattcgtgcttcg 1260
 L  N  V  S  S  L  P  H  A  L  F  M  G  W  L  F  I  R  A  S  161
tacattgcgcctgcgctccagaaggtgacagacttctgcatctggttgttccttcttcag 1320
 Y  I  A  P  A  L  Q  K  V  T  D  F  C  I  W  L  F  L  L  Q  181
tccctggatcgaatcatcttgtgcatgggcgctgtgtatatcaagtggaagaaaatcaag 1380
 S  L  D  R  I  I  L  C  M  G  A  V  Y  I  K  W  K  K  I  K  201
cctacgcctgtgagtccctccctggagtctgatgacatcgagcaaccggataaaggccac 1440
 P  T  P  V  S  P  S  L  E  S  D  D  I  E  Q  P  D  K  G  H  221
cccatgtgcttggtgcaaatacccatgtgcaacgagcgggagtgttacgagcaatcgatt 1500
 P  M  C  L  V  Q  I  P  M  C  N  E  R  E  C  Y  E  Q  S  I  241
tcagcagtttgccagcttgactggccgaagaaccggatcttgatccaggtgctggatgac 1560
 S  A  V  C  Q  L  D  W  P  K  N  R  I  L  I  Q  V  L  D  D  261
tcatctgatgaggaggtggcggggctgatcgaaacagaagtgaagaaatggcagcaaaaa 1620
 S  S  D  E  E  V  A  G  L  I  E  T  E  V  K  K  W  Q  Q  K  281
ggaatcaacataatctacaggcatcggactgaccgtaccggttacaaggctggaaacatg 1680
 G  I  N  I  I  Y  R  H  R  T  D  R  T  G  Y  K  A  G  N  M  301
aaggcgggtatggaatgtgactacgtcaaggattacgagtttgtggccatcttcgatgcc 1740
 K  A  G  M  E  C  D  Y  V  K  D  Y  E  F  V  A  I  F  D  A  321
gactttcaaccaaaatctgactttctcaagcttacaatcccccatttcaaggacaatccg 1800
 D  F  Q  P  K  S  D  F  L  K  L  T  I  P  H  F  K  D  N  P  341
gagttaggccttgtgcaggcccgatgggcctttgtgaataaggatgagaatctgctgaca 1860
 E  L  G  L  V  Q  A  R  W  A  F  V  N  K  D  E  N  L  L  T  361
cggcttcagaatatcaacttgtcctttcacttcgaagttgagcagcaagtgaatggagtg 1920
 R  L  Q  N  I  N  L  S  F  H  F  E  V  E  Q  Q  V  N  G  V  381
```

```
tttctcaatttttttggcttcaatggcaccgcgggcgtgtggcgaatcaaagcactggaa 1980
 F  L  N  F  F  G  F  N  G  T  A  G  V  W  R  I  K  A  L  E  401
gattctggtggttggttggacaggaccactgtggaggatatggacatcgctgttagggct 2040
 D  S  G  G  W  L  D  R  T  T  V  E  D  M  D  I  A  V  R  A  421
catctcaaaggctggaaattcatctttctgaacgatgttcggtccctgtgtgagcttcct 2100
 H  L  K  G  W  K  F  I  F  L  N  D  V  R  S  L  C  E  L  P  441
gagtcgtatgaagcgtaccgcaagcagcaacatcggtggcattctggacctatgcaactt 2160
 E  S  Y  E  A  Y  R  K  Q  Q  H  R  W  H  S  G  P  M  Q  L  461
ttccgtctagctctgccagacattatcaacgctcagatttcatggtcgaagaaatttaac 2220
 F  R  L  A  L  P  D  I  I  N  A  Q  I  S  W  S  K  K  F  N  481
atgatctttctcttctttctcttgagaaagctcatcctaccatttactcctttacatta 2280
 M  I  F  L  F  F  L  L  R  K  L  I  L  P  F  Y  S  F  T  L  501
ttctgcatcatcctacctatgaccatgttcgtcccggaagcaacctaccagcttgggtt 2340
 F  C  I  I  L  P  M  T  M  F  V  P  E  A  T  L  P  A  W  V  521
gtgtgctacattcctgccctcatgtcactcctgaacgtcatcccgtcaccgaagtccttt 2400
 V  C  Y  I  P  A  L  M  S  L  L  N  V  I  P  S  P  K  S  F  541
cctttcctcgtacctttctgctattcgagaacaccatgtccgtcactaagttcaacgcc 2460
 P  F  L  V  P  F  L  L  F  E  N  T  M  S  V  T  K  F  N  A  561
atgatatcgggtctcttccaactgcgtagttcacacgagtgggttgtcaccaagaagtca 2520
 M  I  S  G  L  F  Q  L  R  S  S  H  E  W  V  V  T  K  K  S  581
gggtcaaaagggttagcggacatcgctccatcgtcaacagaggcagagcttttagacgag 2580
 G  S  K  G  L  A  D  I  A  P  S  S  T  E  A  E  L  L  D  E  601
gtcaaggaagtgaagaccgcacctgttgttttgggacgaggtttttcggagtctggtata 2640
 V  K  E  V  K  T  A  P  V  V  L  G  R  G  F  S  E  S  G  I  621
gatacgttgaaacaacaagctgaatcaagcgctccagcgccagtgccgaagaagaaagga 2700
 D  T  L  K  Q  Q  A  E  S  S  A  P  A  P  V  P  K  K  K  G  641
agccgtttgtatcgcaaggagcttactttgtctttcctactcttgactgctgcgggcagg 2760
 S  R  L  Y  R  K  E  L  T  L  S  F  L  L  L  T  A  A  G  R  661
agtttgctctcagcgcaaggaatccacttttacttccttctctttcaaggtatctcgttt 2820
 S  L  L  S  A  Q  G  I  H  F  Y  F  L  L  F  Q  G  I  S  F  681
cttgtggttggtctagatcttataggagagccaaccggttgatctcaggtattacacgaa 2880
 L  V  V  G  L  D  L  I  G  E  P  T  G  -                    694
agtagttcatttttcaagtctatcgaacttgctctctgtgagaagaatagggctccacc 2940
atgtccagcgaccctcaacaagcaagctgattagaaatctcatgcagcttccttacaccg 3000
aagccttcattgcagaaggatgcaatgctgataaaggcttactgccgtttacagtgattg 3060
ggatggtttgtcttcaagccaatcatctgtaccaggaattagtaacaattcagctcgaga 3120
gcagtactggacaatgagaggttgacgagcttctaattttcaatagaagccacgcacccc 3180
attttctttttacttctagctgaaaagaaaggtcgaaggacttcgattgtgaagagtagt 3240
gcctcccagtgcggttcttgaacacgtcaggtttttttaacctgaaaccaattaaattcat 3330
cacgtaattcttttagtcttagctccacgttcagatgcccctgaagcgtgcactgaagct 3360
tgtgctataagcttctttttgtgcgaacc                               3389
```

Figure 5-1 Nucleotide and amino acid sequences of the full length cDNA of PpCslC4. The reading frame starts at position 778(bold) and stops at position 2862(-). The bold and underlined denotes the conserved regions (D, DXD, D, and QXXRW).

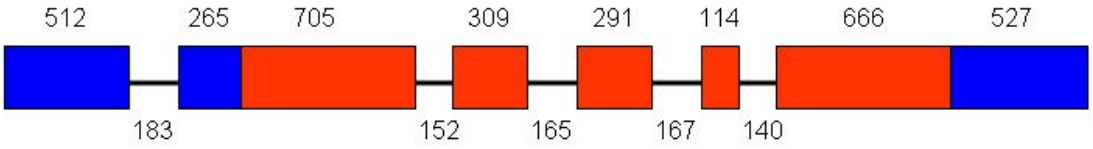
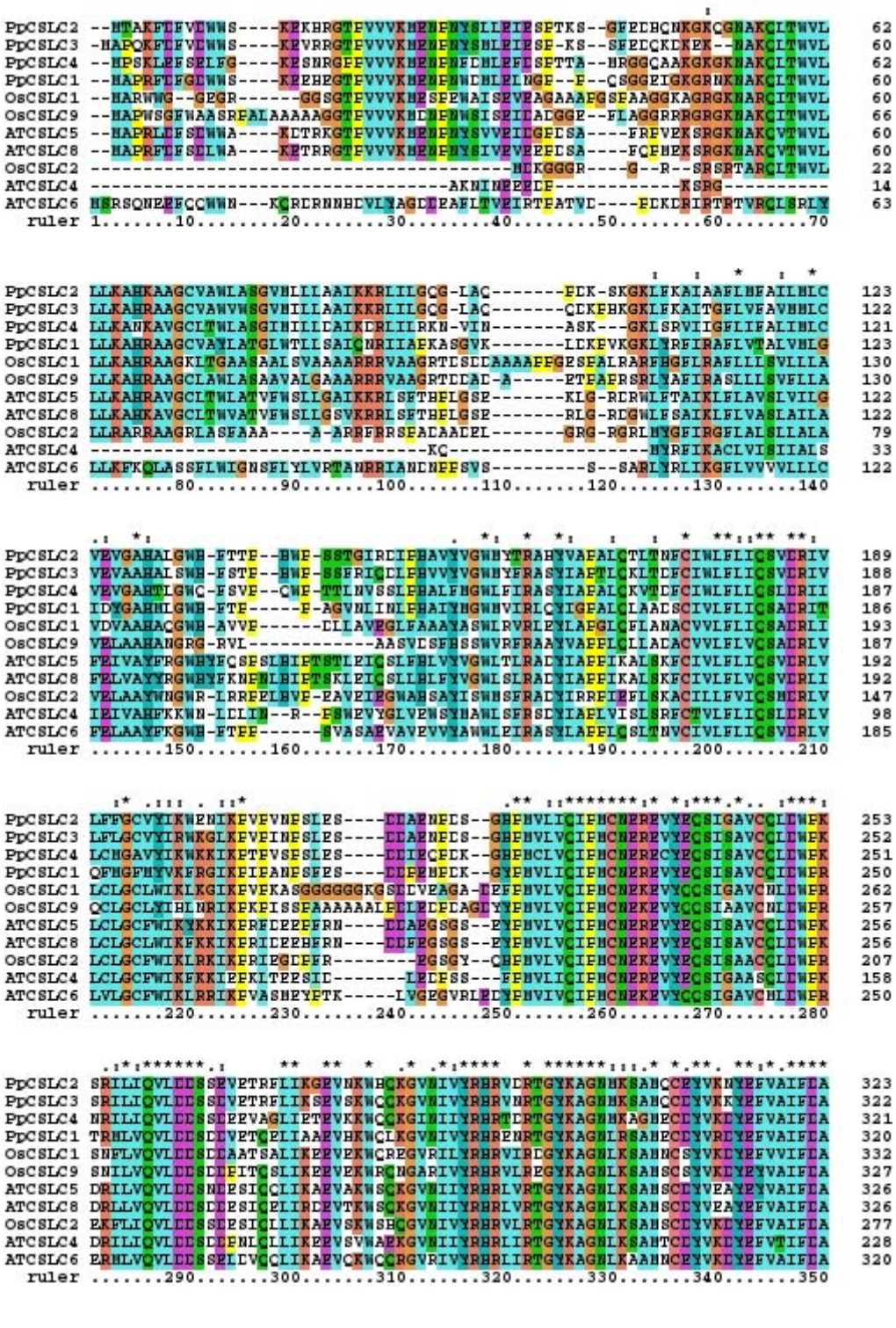
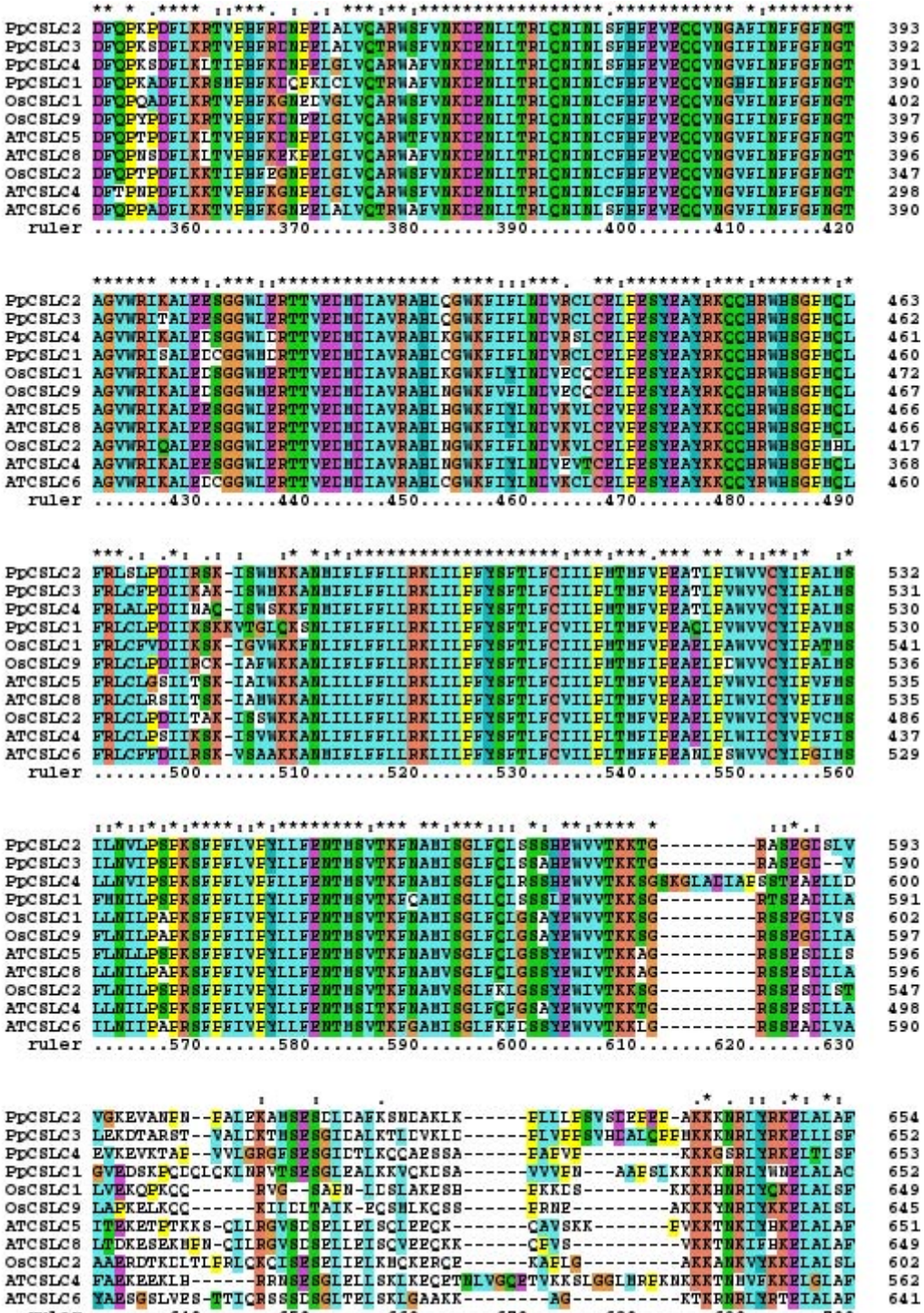


Figure 5-2 gene structure of PpCslC4
Red rectangles represent exons.
Blue rectangles represent UTRs.
Lines represent introns.




PpCSLC2 393
PpCSLC3 392
PpCSLC4 391
PpCSLC1 390
OsCSLC1 402
OsCSLC9 397
ATCSLC5 396
ATCSLC8 396
OsCSLC2 347
ATCSLC4 298
ATCSLC6 390
ruler .....360.....370.....380.....390.....400.....410.....420
PpCSLC2 463
PpCSLC3 462
PpCSLC4 461
PpCSLC1 460
OsCSLC1 472
OsCSLC9 467
ATCSLC5 466
ATCSLC8 466
OsCSLC2 417
ATCSLC4 368
ATCSLC6 460
ruler .....430.....440.....450.....460.....470.....480.....490
PpCSLC2 532
PpCSLC3 531
PpCSLC4 530
PpCSLC1 530
OsCSLC1 541
OsCSLC9 536
ATCSLC5 535
ATCSLC8 535
OsCSLC2 486
ATCSLC4 437
ATCSLC6 529
ruler .....500.....510.....520.....530.....540.....550.....560
PpCSLC2 593
PpCSLC3 590
PpCSLC4 600
PpCSLC1 591
OsCSLC1 602
OsCSLC9 597
ATCSLC5 596
ATCSLC8 596
OsCSLC2 547
ATCSLC4 498
ATCSLC6 590
ruler .....570.....580.....590.....600.....610.....620.....630
PpCSLC2 654
PpCSLC3 652
PpCSLC4 653
PpCSLC1 652
OsCSLC1 649
OsCSLC9 645
ATCSLC5 651
ATCSLC8 649
OsCSLC2 602
ATCSLC4 562
ATCSLC6 641
ruler .....640.....650.....660.....670.....680.....690.....700

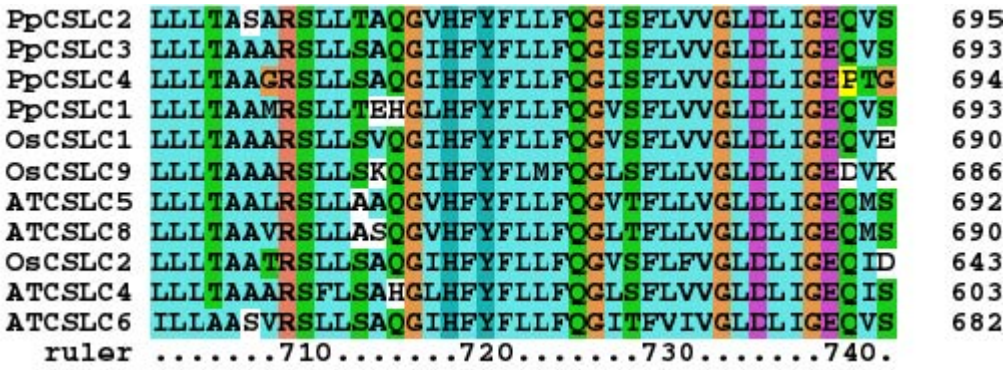


Figure 5-3 Multiple alignments of the deduced amino acids of PpCslC4 with other *Physcomitrella* CslC proteins and *Arabidopsis* and *Oryza* CslCs

PpCSLC1, PpCSLC2, PpCSLC3, PpCSLC4= *Physcomitrella patens* CslC1, CslC2, CslC3, CslC4 (GenBank accession number: ABI55233, ABI55234, ABI55235, ABR26637)

AtCSLC4, AtCSLC5, AtCSLC6, AtCSLC8= *Arabidopsis thaliana* CslC4, CslC5, CslC6, CslC8 (GenBank accession number: NP_566835, NP_194887, NP_187839, NP_180039)

OsCSLC1, OSCSLC2, OSCSLC9=*Oryza sativa* CslC1, CslC2, CslC9 (GenBank accession number: DAA01749, BAD33628, BAC56816)

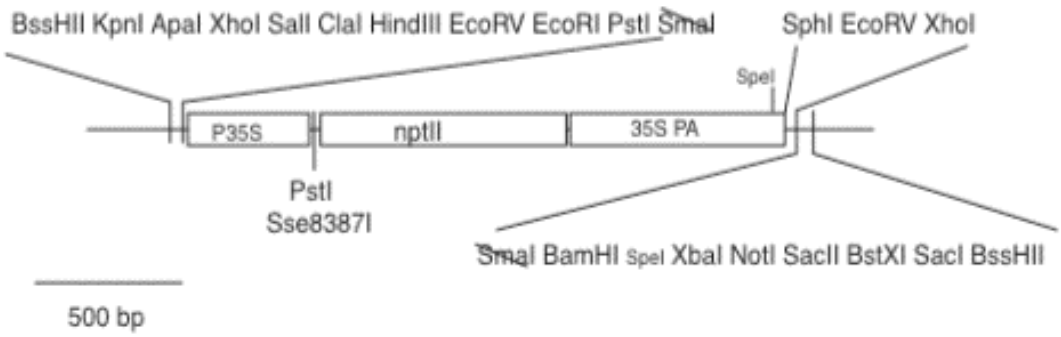


Figure 5-4 pTN3 vector

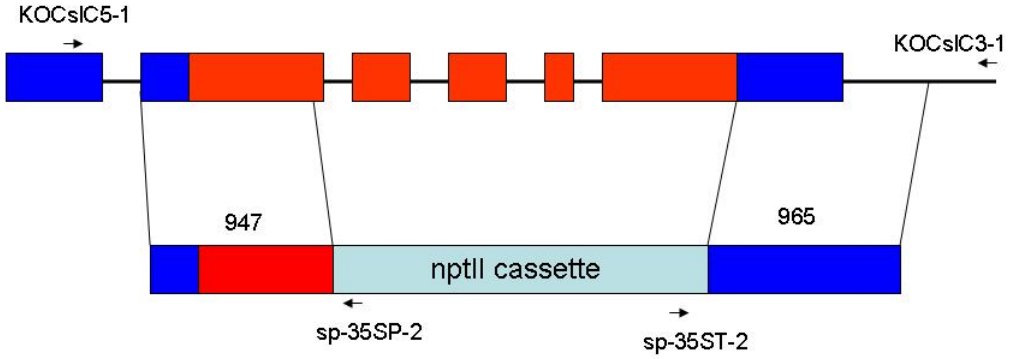


Figure 5-5 Schematic representation of the PpCslC4 gene and KOCslC4 knockout construct

(K) Structure of PpCslC4 gene from *P. Patens.* The red rectangles represent exons, blue rectangles represent UTR, lines represent introns.

(L) The KOCslC4 knockout construct. The 947bp fragment was cloned between the KpnI and ApaI sites of pTN3. The 965bp fragment was cloned between the BamHI and NotI sites of pTNs. The positions of the PCR primers are indicated by arrows.

Primers KOCslC-5-1 and sp-35SP-2, sp-35ST-2 and KOCslC-3-1 were used together, respectively.

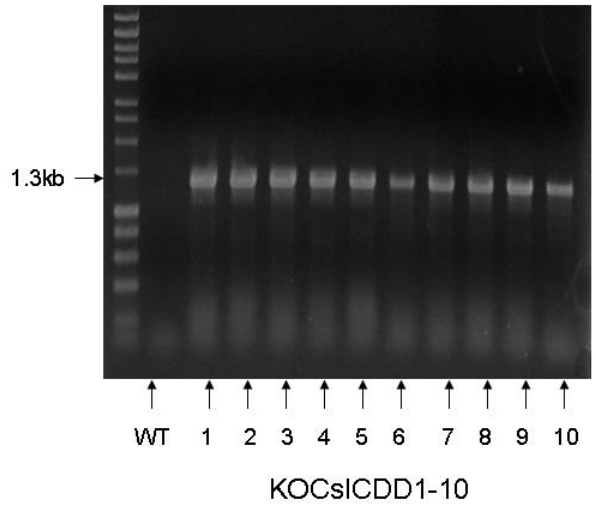


Figure 5-6 PCR analysis of PpCslC4 transformant lines

The selection of 10 transformant lines are shown here.
Primers KOCslC-5-1 and sp-35SP-2 were used in PCR.

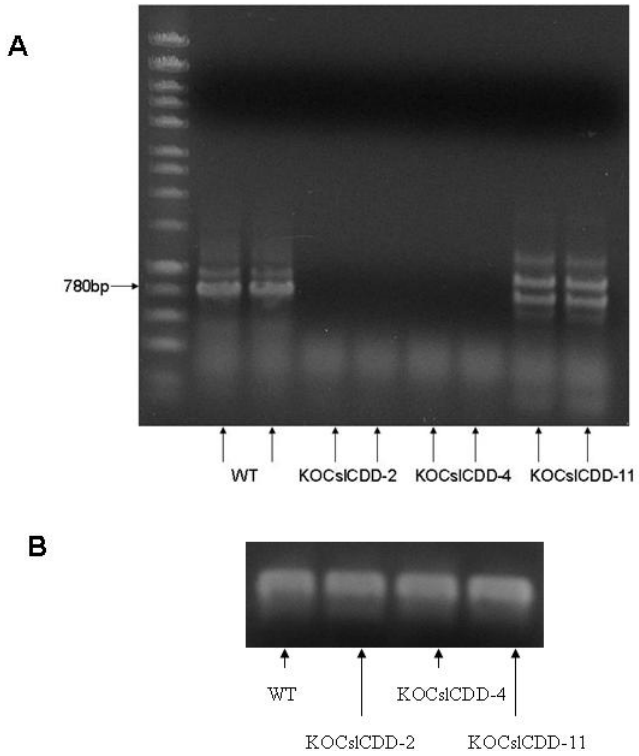


Figure 5-7 RT-PCR analysis of transformant lines

A RT-PCR analysis by using primers RT-CslC-F1 and RT-CslC-R1
Note that the disappearance of product in KOCslCDD-2 and KOCslCDD-4.
Arrow indicates the 780bp product. All other products were caused by mispriming.

B Actin gene of *Physcomitrella* as control

Bibliography

Appenzeller L, Doblin M, Barreiro R, Wang HY, Niu XM, Kollipara K, Carrigan L, Tomes D, Chapman M, Dhugga K. (2004). Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. Cellulose.11: 287–99.

Arioli T, Peng L, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Hofte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE. (1998) Molecular analysis of cellulose biosynthesis in Arabidopsis. Science.279: 717-20.

Ashton NW, Cove DJ. (1977). The isolation and preliminary characterization of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. Mol. Gen. Genet. 154: 87-95.

Ashton NW, Champagne CEM, Weiler T, Verkoczy LK. (2000) The bryophyte *Physcomitrella patens* replicates extrachromosomal transgenic elements. New Phytol. 146: 391–402.

Baskin TI. (2001) On the alignment of cellulose microfibrils by cortical microtubules: a review and a model. Protoplasma. 215: 150-171

Blanton RL, Fuller D, Iranfar N, Grimson MJ, Loomis WF. (2000) The cellulose synthase gene of *Dictyostelium*. Proc Natl Acad Sci USA. 97: 2391-96.

Bowers JE, Chapman BA, Rong J, Paterson AH. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature**.** 422: 433-38.

Britt AB, May GD. (2003) Re-engineering plant gene targeting. Trends plant Sci. 8: 90-95.

Brown RM Jr., Montezinos D. (1976) Cellulose microfibrils: visualization of biosynthetic and orienting complexes in association with the plasma membrane. Proc Natl Acad Sci USA. 73: 143-47.

Brown RM Jr., Willison JHM, Richardson CL. (1976) Cellulose biosynthesis in *Acetobater xylium*: visualization of the site of synthesis and direct measurement of the in vivo process. Proc Natl Acad Sci USA. 73: 4565-69.

Brown RM Jr. (1985) Cellulose microfibril assembly and orientation: recent developments. J. Cell Sci Suppl 2: 13-32.

Brown RM Jr. (1990) Algae as tools in studying the biosynthesis of cellulose: nature's most abundant macromolecule. Experimental phycology: cell walls and surfaces, reproduction, photosynthesis. Eds. Wiessner W, Robinson DG, Starr RC. Springer Verlag. pp19-39.

Brown RM Jr., Saxena IM. (2000) Cellulose biosynthesis: A model for understanding the assembly of biopolymers. Plant Physiol. Biochem. 38: 57-67.

Brown RM Jr. (2004) Cellulose structure and biosynthesis: what is in store for the 21st century? Journal of polymer science. Part A. Polymer chemistry. 42: 487-95.

Burn JE, Hocart CH, Birch RJ, Cork AC, Williamson RE. (2002) Functional analysis of the cellulose synthase genes CesA1, CesA2, and CesA3 in *Arabidopsis*. Plant Physiol. 129: 797-807.

Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shieley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB. (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3; 1,4)-beta-D-glucans. Science.11: 1940-42.

Carpita NC, Gibeaut DM. (1993) Structural models of the primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J. 3: 1-30.

Carpita N, McCann M. (2000) The cell wall. In: Biochemistry and Molecular Biology of Plants. Eds. Buchanan BB, Wihelm G, Johnes RL. American Society of Plant Physiology. pp52-108.

Cocuron JC, Lerouxel O, Drakakaki G, Alonso AP, Liepman AH, Keegstra K, Raikhel N,Wilkerson CG. (2007) A gene from the cellulose synthase-like C family encodes a β-1, 4 glucan synthase. Proc Natl Acad Sci USA.104: 8550-55.

Cove DJ. (1992) Regulation of development in the moss, *Physcomitrella patens*. In: Developmental Biology. A Molecular Genetic Approach. Eds. Brody S, Cove DJ, Ottolengh S, Rosso VEA. Springer Verlag, Heidelberg. pp179-193.

Cove DJ, Knight CD, Lamparter T. (1997) Mosses as model systems. Trends Plant Sci. 2: 99-105.

Cutler S, Somerville C. (1997) Cellulose synthase: cloning by in *silico*. Curr Biol. 7: R108-R111.

Delmer DP. (1999) Cellulose biosynthesis: Exciting times for a difficult field of study. Annu Rev Plant Physiol Plant Mol Biol. 50: 245-76.

Desprez T, Vernhettes S, Fagard M, Refregier G, Desnos T, Aletti E, Pelletier N, Py S, Hofte H. (2002) Resistance against herbicide isoxaben and cellulose deficiency caused by distinct mutations in same cellulose synthase isoform *CESA6*. Plant Physiol. 128: 482–90.

Desprez T, Juraniec M, Crowell EF, Jouy H, Pochylova Z, Parcy F, Hofte H, Gonneau M, Vernhettes S. (2007) Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 104: 15572-77.

Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S, Anderson P. (2004) Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. Science. 303: 363-66.

Doonan JH, Cove DJ. (1985) Immunoflourescence microscopy of mocrotubules in intact cell lineages of the moss, *Physcomitrella patens*. I .Normal and CIPC-treated tip cells. J. Cell Sci. 75: 131-147.

Fagard M, Desnos T, Desprez T., Goubet F, Refregier G, Mouille G, McCann M, Rayon C, Vernhettes S, Hofte H. (2001) PROCUSTE encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of *Arabidopsis*. Plant Cell. 12: 2409-24.

Favery B, Ryan E, Foreman J, Linstead P, Boudonck K, Steer M, Shaw P, Dolan L. (2001) KOJAK encodes a cellulose synthase-like protein required for root hair cell morphogenesis in *Arabidopsis*. Genes Dev. 15: 79–89.

Franz G, Blaschek W. (1990) Cellulose. In: Methods in Plants Biochemistry. Vol II Carbohydrates. Eds. Dey PM, Harborne JB. Academic Press. pp 291-322.

French AD, Miller DP, Aabloo A. (1993) Miniature crystal models of cellulose polymorphs and other carbohydrates. Int. J. Biol Macromol. 15: 30-36.

Griffin D, Buck WR. (1989) Taxonomic and phylogenetic studies on the Bartramiacea. Bryologist. 92: 368-80.

Grimsley NH, Ashton NW, Cove DJ. (1977) The production of somatic hybrids by protoplast fusion in the moss, *Physcomitrella patens*. Mole Gen Genet. 154: 97-100.

Hamann T, Osborne E, Youngs HL, Mission J, Nussaume L, Somerville C. (2004) Global expression analysis of CESA and CSL genes in *Arabidopsis*. Cellulose. 11: 279-286.

Hanin M, Volrath S, Bogucki A, Briker M, Ward E, Paszkowski J. (2001) Gene targeting in *Arabidopsis*. Plant J. 28: 671–77.

Hanin M, Paszkowski J. (2003) Plant genome modification by homologous recombination. Curr. Opin. Plant Biol. 6: 157–62.

Hara K, Morita M, Takahashi R, Sugita M, Kato S, Aoki S. (2001) Characterization of two genes, Sig1 and Sig2, encoding distinct plastid σ factor in the moss *Physcomitrella patens*: phylogenetic relationships to plastid σ factors in higher plants. FEBS Lett. 499: 87-91.

Hazen SP, Scott-Craig JS, Walton JD. (2002) Cellulose synthase-like genes of rice. Plant Physiol. 128: 336-40.

Hebant C. (1977) The Conducting Tissues of Bryophytes. J. Cramer, Vaduz, Liechtenstein, pp157.

Hedenas L. (1990) Axillary hairs in pleurocarpous mosses-a comparative study. Lindbergia. 15: 166-80.

Hohe A, Rensing SA, Mildner M, Lang D, Reski R. (2002) Day length and temperature strongly influence sexual reproduction and expression of a novel MADS-box gene in the moss *Physcomitrella patens*. Plant Biol. 4: 595-602.

Hohe A, Egener T, Lucht JM, Holtorf H, Reinhard C, Schween G, Reski R. (2004) An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, *Physcomitrella patens.* Curr Genet. 44: 339-47.

Holland N, Holland D, Helentjaris T, Dhugga KS, Xoconostle-Cazares B, Delmer DP. (2000). A comparative analysis of the plant cellulose synthase (CesA) gene family. Plant Physiol. 123: 1313–23.

Joshi CP, Bhandari S, Ranjan P, Kalluri UC, Liang X, Fujino T, Samuga A. (2004) Genomics of cellulose biosynthesis in poplars. New Phytol. 164: 53–61.

Kamisugi Y, Cuming AC, Cove DJ. (2005) Parameters determining the efficiency of gene targeting in the moss *Physcomitrella patens*. Nucleic Acids Res. 33: e173.

Kamisugi Y, Schlink K, Rensing SA, Schween G, Stacklberg MV, Cuming AC, Reisk R, Cove DJ. (2006) The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. Nucleic Acids Res. 34: 6205–14.

Kim CM, Park SH, Je BI, Park SH, Park SJ, Piao HL, Eun MY, Dolan L, Han C. (2007) *OsCSLD1*, a cellulose synthase-like D1 gene, is required for root hair morphogenesis in rice. Plant Physiol. 143: 1220-30.

Kimura S, Laosinchai W, Itoh T, Cui XJ, Linder CR, Brown RM. (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis* . Plant Cell. 11: 2075-85.

Knight CD, Cove DJ, Cuming AC, Quatrano RS. (2002) Moss gene technology. In: Molecular plant Biology Volume Two. Eds: Gilmartin PM and Bowler C. Oxford University Press. pp285-301.

Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, Delmer D. (2002) Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. Proc Natl Acad Sci USA. 99: 11109-14.

Liepman AH, Wilkerson CG, Keegstra K. (2005) Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. Proc Natl Acad Sci USA. 102: 2221-26.

Liepman AH, Nairn CJ, Willats WG, Sørensen I, Roberts AW, Keegstra K. (2007) Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. Plant Physiol. 143: 1881-93.

Ligrone R. (1986) Structure, development and cytochemistry of mucilage-secreting hairs in the moss *Timmella barbuloides* (Brid.) Moenk. Ann Bot (London). 58: 859-68.

Lynch M, Force A. (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics. 154: 459–73.

Matthysse AG, Deschet K, Williams M, Marry M, White AR, and Smith WC. (2004) A functional cellulose synthase from *ascidian* epidermis. Proc Natl Acad Sci USA. 101: 986-91.

Muller SC, Brown RM Jr. (1980) Evidence for an intramembranous component associated with a cellulose microfibril synthesizing complex in higher plants. J. Cell Biol. 84: 315-16.

Nishiyama T, Hiwatashi Y, Sakakibara K, Kato M, Hasebe M. (2000) Tagged mutagenesis and gene-trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. DNA Res. 7: 1-9.

Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, and Hasebe M. (2003). Comparative genomics of *Physcomitrella patens* gemetophytic transcriptome and *Arabidopsis thaliana*: Implication for land plant evolution. Proc Natl Acad Sci. USA. 100: 8007-12.

Nishiyama Y, Sugiyama J, Chanzy H, Langan P. (2003) Crystal structure and hydrogen bonding system in cellulose I_α from synchrotron X-ray and neutron fiber diffraction. J. Am Chem Soc. 125: 14300-06.

Nobles D, Brown RM Jr. (2004) The pivotal role of cyanobacteria in the evolution of cellulose synthases and cellulose synthase-like proteins. Cellulose. 11: 437-48.

Nowark MA, Boerlijst MC, Cooke J, Smith JM. (1997) Evolution of genetic redundancy. Nature. 388: 167-71.

Ojangu EL, Jarve K, Paves H, Truve E. (2007) *Arabidopsis thaliana* myosin XIK is involved in root hair as well as trichome morphogenesis on stems and leaves. Protoplasma. 230: 193-202.

Paradez AR, Somerville CR, Ehrhardt DW. (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. Science. 312: 1491-95.

Parihar NS. (1961) An introduction to embryophyte. Vol I. Bryophyta. Central Book Depot. Allahabad. pp1-6.

Paterson AH, Bowers JE, Chapman BA. (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA.101: 9903-08.

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM. (1996) Higher plants contain homologs of the bacterial *cel*A genes encoding the catalytic sun-unit of cellulose synthase. Proc Natl Acad Sci USA. 93: 12637-42.

Persson S, Paredez A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR. (2007) Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. Proc Natl Acad Sci USA. 104: 15566-71.

Pesch M, Hulskamp M. (**2**004) Creating a two-dimensional pattern de novo during *Arabidopsis* trichome and root hair initiation. Curr Opin Genet Dev. 14: 422-27.

Preston RD. (1974) Cellulose. In: The Physical Biology of Plant Cell Walls. Chapman and Hall, London. pp444-56.

Quatrano R, Bashiardes S, Cove D, Cuming A, Knight C, Clifton S, Marra M, Hillier L, Pape D, Martin J. (1999) Leeds/Wash U. Moss EST Project. http://www.ncbi.hlm.nih.gov or http://www.moss.leeds.ac.uk.

Quatrano RS, McDaniel SF, Khandelwal A, Perroud PF, Cove DJ. (2007) *Physcomitrella patens*: mosses enter the genomic age. Curr Opin Plant Biol. 10: 182-89.

Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, Van de Peer Y, Reski R. (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. BMC Evol Biol. 7: 130.

Reski R. (1998) Development, genetics and molecular biology of mosses. Bot. Acta. 111: 1-15.

Richmond T. (2000) Higher plant cellulose synthases. Genome Biol. 4: 30011–16.

Richmond TA, Somerville CR. (2000). The cellulose synthase superfamily. Plant Physiol. 124: 495-98.

Richmond T, Somerville CR. (2001) Integrative approaches to determining Csl function. Plant Mol Biol. 47: 131-43.

Roberts AW, Roberts E. (2004) Cellulose synthase (CesA) genes in algae and seedless plants. Cellulose. 11: 419-35.

Roberts A, Bushoven JT. (2007) The cellulose synthase (CesA) gene superfamily of the moss *Physcomitrella patens*. Plant Mol Biol. 63: 207-19.

Roelofsen S. (1958) Cell wall structure as related to surface growth. Acta Botanica Neerlandica. 7: 77-89.

Ross P, Mayer R, Benziman M. (1991) Cellulose biosynthesis and function in bacteria. Microbiol Rev. 55: 35-58.

Sakakibara K, Nishiyama T, Sumikawa N, Kofuji R, Murata T, Hasebe M. (2003) Involvement of auxin and a homeodomain-leucine zipper I gene in rhizoid development of the moss *Physcomitrella patens*. Development. 130: 4835-46.

Saxena IM, Lin FC, Brown RM Jr. (1990) Cloning and sequencing of the cellulose synthase catalytic sub-unit gene of *Acetobacter xylium*. Plant Mol Biol. 15: 673-83.

Saxena IM, Brown RM Jr., Fevre M, Geremia RA, Hernissat B. (1995) Multidomain architecture of β-glycosyl transferase: implications for mechanism of action. J. Bacteriol. 177: 1419-24.

Saxena IM, Brown RM Jr. (1997). Identification of cellulose synthase(s) in higher plants: Sequence analysis of processive ß-glycosyltransferases with the common motif "D, D, D35Q(R, Q)XRW." Cellulose. 4**:** 33-49.

Schaefer D, Zryd JP, Knight C, Cove DJ. (1991) Stable transformation of the moss *Physcomitrella patens*. Mol Gen Genet. 226: 418–24.

Schaefer DG, Zryd JP. (1997) Efficient gene targeting in the moss *Physcomitrella patens*. Plant J. 11: 1195–1206.

Schaefer DG. (2001) Gene targeting in *Physcomitrella patens*. Curr Opin Plant Biol. 4: 143-150.

Schaefer DG, Zryd JP. (2001) The moss *Physcomitrella patens*, Now and Then. Plant Physiol. 127: 1430-38.

Schaefer DG. (2002) A new moss genetics: targeted mutagenesis in *Physcomitrella patens*. Annu Rev Plant Biol. 53: 477-501.

Schween G, Gorr G, Hohe A, Reski R. (2003) Unique tissue-specific cell cycle in *Physcomitrella*. Plant Biol. 5: 1-9.

Shaw AJ, Goffinet B. (2000) Bryophyte Biology. Cambridge University Press. pp76-77.

Smith GM. (1955) Cryptogamic Botany, Vol II. Bryophytes and Pteridophytes. 2nd Ed. McGraw-Hill, New York.

Somerville C. (2006) Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol. 22: 53-78.

Sugiyama J, Vuong R, Chanzy H. (1991) Electron diffraction study of two crystalline phases occurring in native cellulose from an algal cell wall. Macromolecules. 24: 4168-75.

Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H. (2003) Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthase in the secondary wall. Plant Physiol. 133: 73-78.

Taylor NG, Laurie S, Turner SR. (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell.12: 2529-40.

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR. (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci USA. 100: 1450–55.

Taylor NG. (2007) Identification of cellulose synthase AtCesA7 (IRX3) in vivo phosphorylation sites-a potential role in regulating protein degradation. Plant Mol Biol. 64: 161-71.

The *Arabidopsis* genome initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis* thaliana. Nature. 408: 796-815.

Thiessen G, Munster T, Henschel K. (2001) Why don't mosses flower? New Phyto. 150: 1-5.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-82.

Tseko I. (1999) The sites of cellulose synthesis in algae: diversity and evolution of cellulose-synthesizing enzyme complexes. J Phycol. 35: 635-55.

Tusnády GE and Simon I. (1998) Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. J. Mol. Biol. 283: 489-506.

Tusnády GE and Simon I. (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics. 17: 849-50.

Vergara CE, Carpita NC. (2001) beta-D-glycan synthase and the CesA gene family: lessons to be learned from the mixed-linkage (1→3), (1→4) –D-glucan synthase. Plant Mol Biol. 47: 145-60.

Wang X, Cnops G, Vanderhaeghen R, Block SD, Montagu MV, Van Lijsebettens M. (2001) AtCSLD3, a cellulose synthase-like gene important for root hair growth in *Arabidopsis*. Plant Physiol. 126: 575–86.

Zhang JZ. (2003) Evolution by gene duplication: an update. Trends Ecol Evol. 18: 292-98.

Vita

Hua Zhang Wise was born in Tanghe, Henan China on August 2, 1972, the daughter of Qingshan Zhang and Guifen Li. She graduated from the First High School of Tanghe in 1988. She received a Bachelor of Science in Biology from XinYang Teacher's University, China in 1992 and a Master of Science in Botany from Wuhan University, China in 1995. In the following 5 years she was employed as a research scientist at the Research and Development Center, Changsha Tobacco Company, China. In September 2000, she entered the Graduate School of the University of Texas at Austin.

Permanent Address: 8206 Dixon Drive, Austin TX 78745

This dissertation was typed by the author.